TEEN DATING VIOLENCE: MEASUREMENT AND OUTCOMES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Deinera Marea Exner

January 2014

TEEN DATING VIOLENCE: MEASUREMENT AND OUTCOMES

Deinera Marea Exner, Ph. D.

Cornell University 2014

Abstract

Violence experienced in early and mid-adolescent romantic relationships (known as *teen dating violence*) is an important public health issue, and the three papers in this volume each address a different research question on this topic. Emerging research demonstrates that individuals who experience victimization in adolescence are more likely to be re-victimized in future relationships; however, past work on this topic is limited by potential confounding, and lack of assessment of potential mediators of this relationship. Thus, the first paper (Chapter Two) used data from the National Longitudinal Study of Adolescent Health to explore pathways to re-victimization, adjusting for confounding using a high-dimension propensity score. Results indicated that dating violence experienced during adolescence was indirectly associated with intimate partner violence experienced 12 years later, through the experience of intimate partner violence at 5.5 year follow-up.

These findings, as well as all empirical findings in the field, rest on the quality of measurement, and so the selection of a measure for a given research study is an important task. Currently, however, no comprehensive compendium exists that presents teen dating violence measures with evidence of reliability and validity and discusses strengths and limitations of these evidence-based measures. Thus, the second paper (Chapters Three and Four) presents a two-part comprehensive review of teen dating violence measures that have been the focus of

psychometric testing. This review also summarizes empirical literature that uses identified measures.

Due to the complex and nuanced nature of interpersonal interactions, psychological aggression is a particularly difficult construct to measure. Empirical data show that psychological aggression is common in teen dating relationships, but to more precisely answer questions about the impact of this aggression on healthy development, measures must be designed that capture psychological aggression that is purposeful, serious and perceived as harmful. The third and final paper in this volume (Chapter Five) reports on the initial adaptation of a measure of severe psychological aggression (the Measure of Psychologically Abusive Behaviors; Follingstad, 2011, *Journal of Interpersonal Violence*, *26*(6)) for teen dating relationships. Together, these three papers advance understanding of teen dating violence and support its developmental and public health importance.

BIOGRAPHICAL SKETCH

Deinera Exner (a.k.a. Exner-Cortens) holds a Master of Arts from Cornell University in Developmental Psychology (2011), a Master of Public Health in Social and Behavioral Science from Boston University (2009), and a Bachelor of Science in Cellular, Molecular and Microbial Biology from the University of Calgary (2007). Deinera's research focuses on understanding interpersonal violence in intimate relationships. Past and current projects in this area include media framing of domestic homicides in Botswana, intimate partner violence in the lives of Canadian Aboriginal women, prevalence of sexual violence in gay, lesbian and bisexual populations in the United States, longitudinal outcomes of teen dating violence victimization, the evaluation of a campus-based sexual violence prevention program and qualitative research on the use of orders of protection by teens in New York State. Upon the completion of her doctoral degree, Deinera will be a post-doctoral fellow at the Centre for Addiction and Mental Health's Centre for Prevention Science in London, Ontario, Canada.

To my family

ACKNOWLEDGMENTS

This dissertation reflects the mentorship, guidance and support of numerous individuals. Writing this document, it became very clear to me how much each person has contributed to my development as a critical thinker, researcher and person, and for that, I am incredibly grateful. First, thank you to my chair, John Eckenrode, for your constant support and quiet guidance. I have gained so much from working with you, both personally and professionally, and am very appreciative of all your help and mentorship. To my minor members, John Bunge, Dawn Schrader and Emily Rothman; each of you has played a different role in my development, and I appreciate each and every contribution you have all made. To my other mentors, especially Gary Evans, Anthony Ong and Janis Whitlock (and all the others who played roles big and small), thank you. I learned so much about different approaches and topics working with each of you, and feel very fortunate for the experience. To the staff at the Ithaca Advocacy Center and Nina Cummings at Gannett Health Services, thank you for helping me learn how things work on the ground. I am so appreciative that you took the time to share your thoughts and perspectives with me, and hope I can represent them well in my work. To all my undergraduate research assistants – Martine Aurelien, Rachel Bernstein, Briana Burris, Sade Famakinwa, Natalie Finn, Ryan Hand, Katie Kellner, Juliana Kennedy, Salome Odera, Lauren Ritter, Andrea Sanchez, Erika Shankman and Alex Simson – thank you for your hard work and dedication. Thanks especially to Lydia Gill for serving as lab manager, it was such a pleasure to work with you. To everyone at the Bronfenbrenner Center for Translational Research, your work inspires me to continue to do applied research, and I am more grateful than I could ever say that I had the chance to spend these past four years working and learning at BCTR. To my fellow graduate students at the center and in the department (you know who you are), I am so glad we had each other, and am so appreciative of the support each of you gave to me, especially in those final weeks of writing, when speaking in coherent sentences became daunting! And finally, to my family, who have given me unwavering love and support throughout this whole process. Thank you.

I would also like to thank my funders, specifically the Canadian Institutes of Health

Research, who provided a doctoral fellowship that supported my study for the past two years, and the Department of Human Development, whose dissertation funding and summer stipend allowed me to complete my focus group data collection. Thanks also to Alberta Scholarship Programs, whose Sir James Lougheed Award of Distinction funded my second year of studies, and to the Bronfenbrenner Center for Translational Research, who supported my application for Add Health data.

I could not have collected the focus group data without the assistance of the Bronfenbrenner Center for Translational Research. Thanks especially to Jane Powers and Sara Birnel Henderson, as well as to each of the sites. And to the adolescent participants for talking to me, and teaching me how to make our research better, thank you.

Finally, thanks to my editor. You have been unbelievably supportive, and made every word in this document better. I would not have gotten through this process without you, and am so blessed to have your love and support.

TABLE OF CONTENTS

CHAPTER ONE

TEEN DATING VIOLENCE MEASUREMENT AND OUTCOMES: AN INTRODUCTION

Teen Dating Violence Measurement and Outcomes: An Introduction

Adolescence is a time of social, cognitive and biological transition (Steinberg, 2008a). As part of this transition, adolescents experience psychosocial development, including the development of identity, intimacy, sexuality, autonomy and achievement (Steinberg, 2008a). The attainment of these developmental tasks is supported by a number of interpersonal contexts, including relationships with parents, peers and romantic partners. Until recently, the latter context was considered trivial and was not the focus of much research attention (Brown, Feiring, & Furman, 1999; Collins, 2003). However, adolescent romantic relationships are common (Carver, Joyner, & Udry, 2003), and are theoretically implicated in the development of all these psychosocial tasks (Brown et al., 1999; Collins & Sroufe, 1999; Collins, Welsh, & Furman, 2009; Furman & Shaffer, 2006); participation in adolescent dating relationships is also related to relationship outcomes and quality in young adulthood (Joyner & Campa, 2006; Meier & Allen, 2009).

While romantic relationships can play a positive role in youth development (Furman, Low, & Ho, 2009), Roisman, Masten, Coatsworth and Tellegen (2004) suggest that adolescent romantic relationships are an emerging, and not salient, developmental task. Serious, off-time engagement in these relationships may interrupt the accomplishment of more salient tasks (i.e., friendship, academic and conduct), and empirical work has demonstrated that associations between participation in adolescent romantic relationships and negative behaviors may be especially pronounced for younger adolescents (Neemann, Hubbard, & Masten, 1995). As an emerging task, adolescent romantic relationships serve as an important domain to explore and gain experience (Roisman et al., 2004); one domain for this exploration is conflict negotiation (Selman, 1980), a skill which may be poorer in emerging developmental settings (Mischo, 2005; Selman, 1980). For example, examining conflict negotiation with friends and romantic partners,

Furman and Shomaker (2008) found that mid-adolescent participants exhibited more conflict with romantic partners than friends during a video-taped discussion task, potentially because peer relationships were a more salient, and familiar, interpersonal context. Adolescent romantic relationships are also characterized by heightened emotionality and volatility (Feiring & Furman, 2000; Giordano, Manning, & Longmore, 2006), related to the development of the cognitive control system (Steinberg, 2008b). This system develops over the course of adolescence, with maturation leading to greater inhibition of impulsive behavior and better control over responses to emotionally arousing situations (Steinberg, 2008b). In sum, adolescent romantic relationships are an important but emerging developmental context that can support the accomplishment of psychosocial development. Romantic relationships also serve as a place to develop conflict negotiation skills in a new interpersonal context, one with pronounced emotionality to which early and mid-adolescents have less ability to regulate their response.

Given this developmental background, it is perhaps not surprising that a large number of adolescents report the experience of psychologically aggressive behaviors in their dating relationships (e.g., insulting, yelling, treating disrespectfully); as shown by Exner-Cortens, Gill and Eckenrode (Chapter Three), the number of adolescents reporting these behaviors in quantitative survey data often exceeds 50%, with some studies reporting that almost all participants endorsed the experience of at least one psychologically aggressive behavior. Adolescents also use and experience other unhealthy behaviors in their romantic relationships, including physical violence, sexual violence and stalking (CDC, 2012). When they occur in early- and mid-adolescent romantic relationships, these four behaviors are referred to as teen dating violence, the focus of a growing body of literature (Foshee & Reyes, 2011). Since many adolescents have the capacity to use some of these behaviors, recent work considers dating violence as part of a continuum of healthy and unhealthy relationship behaviors. Reflecting on

this continuum, we characterize healthy relationships as those with infrequent use of low-severity psychologically aggressive behaviors (i.e., predominately healthy in nature); unhealthy relationships by more frequent use of low-severity psychologically aggressive behaviors, and infrequent use of moderate-severity psychologically aggressive and low-severity physically aggressive behaviors (i.e., a mix of healthy and unhealthy); and teen dating violence relationships by the purposeful and consistent use of low, moderate and severe psychologically, physically and/or sexually aggressive behaviors (i.e., predominately unhealthy). These definitions reflect the changing state of empirical understanding and also incorporate a developmental perspective; however, they do not accurately characterize most existing research, or research that is described in several of the papers in this volume, which considers the endorsement of any unhealthy behavior indicative of teen dating violence. Rather, we put forth these definitions to encourage future research designs that more carefully delineate aggressive experiences in romantic relationships, so that we can better understand differences (and similarities) between healthy, unhealthy and violent relationships. Indeed, the goal of this volume is to draw attention to limitations in the field, with each paper aiming to address one limitation, and all papers identifying important future directions for research.

Existing research documents that violent and aggressive behaviors occur in a substantial minority of adolescent romantic relationships.[1] Nationally, approximately 10% of adolescents report that they were hit, slapped or physically hurt on purpose by a boyfriend or girlfriend in the past year, a prevalence that is comparable to other major public health concerns; for example, in this same sample, 12.9% of adolescents reported not using any method to prevent pregnancy at

[1] For clarity, and to reflect prior research findings, we will refer to physically, sexually and psychologically aggressive behaviors used and experienced in adolescent romantic relationships as teen dating violence throughout this volume. Given the state of the field, studies examining teen dating violence likely include individuals from across the healthy-unhealthy relationship spectrum, and so findings should be considered more indicative of behavioral experiences in relationships, as opposed to a type of relationship. Also, because stalking was only recently added to the definition of teen dating violence, past empirical work that discusses teen dating violence is unlikely to include a specific consideration of this behavior, though may consider it as part of psychological aggression.

last sex, 6.4% reported smoking on 20 or more days in the past month and 2.3-6.8% reported other illicit drug use (CDC, 2012). While the experience of psychological and mild physical violence is generally gender-equitable, females are more likely to experience sexual (Wolitzky-Taylor et al., 2008) and severe physical violence (Coker et al., 2000). Teen dating violence victimization is also linked to multiple adverse health outcomes in males and females, including suicidal thoughts and behaviors, substance use and depressive symptomatology (e.g., Ackard, Eisenberg, & Neumark-Sztainer, 2007; Exner-Cortens, Eckenrode, & Rothman, 2013; Roberts, Klein, & Fisher, 2003), as well as to injury (Foshee, 1996) and death (Bureau of Justice Statistics, 2011), with these latter two outcomes disproportionately affecting females. There is also growing evidence that victimization in early and mid-adolescent romantic relationships is associated with re-victimization in late adolescence and young adulthood (e.g., Cui, Gordon, Ueno, & Fincham, 2013; Exner-Cortens et al., 2013; Gómez, 2011; Smith, White, & Holland, 2003; Williams, Connolly, Pepler, Craig, & Laporte, 2008).

Given its prevalence and association with adverse health and well-being, dating violence is a considerable public health problem: associations with future re-victimization are of special concern, as adult intimate partner violence is linked to a number of negative outcomes for both the recipient and children in the household (Bair-Merritt, Blackstone, & Fuedtner, 2006; Campbell, 2002; Coker et al., 2002; Pavao, Alvarez, Baumrind, Induni, & Kimerling, 2007; Staggs & Riger, 2005). While several studies have now explored this association, none of this past work has explored potential mediating mechanisms. Further, research continues to document that individuals who experience teen dating violence differ from their non-victimized peers on a number of dimensions (Knoble, Capaldi, Shortt, & Kim, 2012; Lewis & Fremouw, 2001; Offenhauer & Buchalter, 2011; Sherer, 2009), such that controlling only for socio-demographics, as is common in most empirical studies, may not account for all the potential

confounds of this relationship, potentially resulting in spurious or suppressed associations. To address these issues, Exner-Cortens, Eckenrode, Bunge and Rothman (Chapter Two) explored the association between teen dating violence experienced in middle and high school and intimate partner violence experienced 12 years later, using data from the National Longitudinal Study of Adolescent Health (Add Health). To investigate pathways to re-victimization, this paper considered potential mediation by depressive symptoms, substance use and intimate partner violence experienced at 5.5-year follow-up. All analyses in this paper controlled for a high-dimension propensity score, either by model adjustment or matching followed by model adjustment. The propensity score method is used to meet the assumption of strongly ignorable treatment assignment, by allowing for adjustment on multiple confounders of the predictor-outcome relationship. Results of this study indicated that dating violence experienced during early and mid-adolescence was indirectly associated with intimate partner violence experienced 12 years later, through the experience of intimate partner violence at 5.5 year follow-up.

Like victimization, perpetration of dating violence is fairly common, and begins in middle school with the onset of adolescent dating (Foshee & Reyes, 2007). Similar to what is generally observed for victimization prevalence, Foshee et al. (2009) found that psychologically aggressive behaviors (e.g., insulting in front of others) were used most commonly, followed by moderate physical (e.g., pushing), severe physical (e.g., choking) and sexually aggressive behaviors, with males and females equally perpetrating psychological and moderate physical violence, but with males more likely to report the use of severe physical and sexual violence. Using a random sub-sample of participants from this same study, Foshee, Bauman, Linder, Rice and Wilcher (2007) found that just over half the female participants described that they perpetrated dating violence in self-defense. Over the course of early and mid-adolescence, Foshee et al. (2007) reported that perpetration of each of these behaviors increased for both

males and females, peaking for moderate physical perpetration around age 17 and severe physical and sexual perpetration around age 16, and then desisting; however, the use of psychological aggression continued to increase through age 19, perhaps because of the more normative and/or less sanctioned use of this aggression. The group desistance of the majority of these behaviors is similar to other adolescent-limited antisocial behaviors (Moffitt, 1993), and likely reflects cognitive maturation (Steinberg, 2008b). Risk factors for victimization and perpetration overlap substantially (Exner-Cortens, Eckenrode, Bunge, & Rothman, Chapter Two; Vagi et al., 2013), which may be because individuals are likely to both use and receive these behaviors within an aggressive adolescent relationship (e.g., Giordano, Soto, Manning, & Longmore, 2010; Gray & Foshee, 1997; Miller et al., 2013; Orpinas, Hsieh, Song, Holland, & Nahapetyan, 2013).

However, it must be noted that, while certainly important in defining the broad parameters of teen dating violence, the findings of this past work, including findings presented in the second chapter of this volume, rest on the quality of measurement. While this may seem apparent, it can be forgotten when the focus is on expeditiously understanding and addressing an issue of public health concern. As was discussed previously in this chapter, by considering any endorsement of an aggressive or violent behavior indicative of a violent relationship, current empirical literature may not adequately capture dating violence in a way that matches lived experience; this is especially true for the construct of psychological aggression, which suffers from poor definition and validation (Follingstad, 2007). Further, many teen dating violence measures were developed for adults (Wolfe et al., 2001), and it is not clear which measures have evidence of reliability and validity in early and mid-adolescent populations, and if measures with this evidence are used in empirical research. For example, the questions in Add Health come from the Revised Conflict Tactics Scale (Straus, Hamby, Boney-McCoy, & Sugarman, 1996), a

popular measure for assessing teen dating violence (Capaldi, Knoble, Shortt, & Kim, 2012), but a measure that was developed and tested in adult and college-age samples. Since what is measured determines what is found, it is important to understand the state of teen dating violence measurement, including which measures have psychometric evidence in adolescent populations. However, no compendium of teen dating violence measures currently exists, and so this question is difficult to answer; thus, in a two-part comprehensive review, Exner-Cortens, Gill and Eckenrode (Chapters Three and Four) explored 13 teen dating violence measures that have been the focus of psychometric testing. As part of this review, they also located empirical articles that had used these 13 measures over the past ten years, demonstrating that most teen dating violence research has not used empirically-tested measures. This compendium will aid researchers in selecting evidence-based measures for their empirical work, and will ideally also stimulate needed discussion around teen dating violence definition and measurement.

From the review presented in Chapters Three and Four, it is also apparent that no existing measure of adolescent psychological aggression fully captures the construct, in large part because these measures generally do not probe for context (e.g., severity, frequency, perceived intent, perception); however, since these actions occur in dyadic interpersonal relationships, failure to consider context vastly limits a full understanding of meaning and experience. Further, when considering teen dating violence, literature reviewed at the beginning of this chapter indicates that some of this behavior may be developmentally normative, but using current measures, we cannot determine which of this behavior is normative and which is not, or investigate whether there are any differences between recipients of more normative aggression and recipients of aggression with negative and purposeful intent.

Further, the measures discussed in Exner-Cortens, Gill and Eckenrode (Chapters Three and Four) were mostly developed a decade ago, and during this time, cyber-bullying has become

a concerning and prevalent experience among adolescents (in 2012, 22.1% of females and 10.8% of males in grades 9-12 reported being electronically bullied through e-mail, chat rooms, instant messaging, Web sites or texting in the past year; CDC, 2012). Recent work has also documented the growing use of cyber-bullying, or electronic psychological aggression, in adolescent romantic relationships (Draucker & Martsolf, 2008; Offenhauer & Buchalter, 2011). The use of cyber-bullying in romantic relationships is not a surprising extension, but is concerning given emerging research demonstrating the potentially enhanced adverse effects of peer cyber-bullying, over and above the effects of traditional bullying, on well-being (Bonanno & Hymel, 2013), which may result from cyber-bullying's public nature and potential for bully anonymity (Sticca & Perren, 2013). As romantic relationships are a newer, more fragile context for adolescents than peer relationships (Sullivan, 1953), it is conceivable that cyber-bullying in this context is also as strongly, if not more strongly, related to adverse outcomes than cyber-aggression experienced in peer relationships.

Given these needs (incorporation of context, items on electronic psychological aggression), Exner-Cortens, Eckenrode, Schrader and Rothman (Chapter Five) discuss the development of a new measure of psychological aggression, the Psychological Aggression in Teens Scale (PATS). This scale is an adaptation of the Measure of Psychologically Abusive Behaviors (MPAB; Follingstad, 2011), a measure designed for adults to tap more egregious forms of psychological aggression (i.e., the receipt of intentional behavior with the potential for long-term harm). To adapt this measure for use in adolescent samples, Exner-Cortens, Eckenrode, Schrader and Rothman (Chapter Five) conducted 10 adolescent focus groups across New York State, and used their feedback to revise the adult MPAB questions. The adaptation also added a number of questions on electronic psychological aggression. While this measure is a promising addition, it requires psychometric testing before it can be used with adolescents.

The three papers presented in this volume were each designed to address a specific limitation facing the field of teen dating violence, and all aim to deepen understanding, to demonstrate its public health and developmental importance and to improve future research. However, each is also a preliminary step toward these broad goals, and future work replicating and expanding their findings is important. There are also a number of remaining questions that these chapters don't address, including whether there is empirical support for developmental theory that may explain why some teens use dating violence, and if so, if this suggests more effective ways to prevent engagement in these behaviors. Given the public health importance of dating violence, we feel it is worthy of continued attention and study, and hope the present volume adds to understanding about these experiences.

References

Ackard, D. M., Eisenberg, M. E., & Neumark-Sztainer, D. (2007). Long-term impact of adolescent dating violence on the behavioral and psychological health of male and female youth. *Journal of Pediatrics, 151*, 476-481.

Bair-Merritt, M. H., Blackstone, M., & Feudtner, C. (2006). Physical health outcomes of childhood exposure to intimate partner violence: A systematic review. *Pediatrics, 117*, e278-e290.

Brown, B. B., Feiring, C., & Furman, W. (1999). Missing the love boat: Why researchers have shied away from adolescent romance. In W. Furman, B. B. Brown, & C. Feiring (Eds.), *The development of romantic relationships in adolescence* (pp. 1-18). New York, NY: Cambridge University Press.

Bonanno, R. A., & Hymel, S. (2013). Cyber bullying and internalizing difficulties: Above and beyond the impact of traditional forms of bullying. *Journal of Youth and Adolescence, 42*, 685-697.

Bureau of Justice Statistics. (2011, November). Homicide trends in the United States, 1980-2008. Retrieved from http://www.bjs.gov/content/pub/pdf/htus8008.pdf

Campbell, J. C. (2002). Health consequences of intimate partner violence. *Lancet, 359*, 1331-1336.

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner Abuse, 3*, 231-280.

Carver, K., Joyner, K., & Udry, J. R. (2003). National estimates of adolescent romantic relationships. In P. Florsheim (Ed.), *Adolescent romantic relations and sexual behavior: Theory, research, and practical implications* (pp. 23-56). Mahwah, NJ: Lawrence Erlbaum Associates.

Centers for Disease Control [CDC]. (2012). *Understanding teen dating violence*. Retrieved from http://www.cdc.gov/violenceprevention/pdf/teendatingviolence2012-a.pdf

Coker, A. L., McKeown, R. E., Sanderson, M., Davis, K. E., Valois, R. F., & Huebner, E. S. (2000).

Severe dating violence and quality of life among South Carolina high school students. *American Journal of Preventive Medicine, 19*, 220-227.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., & Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23*, 260-268.

Collins, W. A. (2003). More than myth: The developmental significance of romantic relationships during adolescence. *Journal of Research on Adolescence, 13*, 1-24.

Collins, W. A., & Sroufe, L. A. (1999). Capacity for intimate relationships: A developmental construction. In W. Furman, B. B. Brown, & C. Feiring (Eds.), *The development of romantic relationships in adolescence* (pp. 125-147). New York, NY: Cambridge University Press.

Collins, W. A., Welsh, D. P., & Furman, W. (2009). Adolescent romantic relationships. *Annual Review of Psychology, 60*, 631-652.

Cui, M., Gordon, M., Ueno, K., & Fincham, F. D. (2013). The continuation of intimate partner violence from adolescence to young adulthood. *Journal of Marriage and Family, 75*, 300-313.

Draucker, C. B., & Martsolf, D. S. (2008). The role of electronic communication technology in adolescent dating violence. *Journal of Child and Adolescent Psychiatric Nursing, 23*, 133-142.

Exner-Cortens, D., Eckenrode, J., & Rothman, E. (2013). Longitudinal associations between teen dating violence victimization and adverse health outcomes. *Pediatrics*, *131*, 71-78.

Feiring, C., & Furman, W. C. (2000). When love is just a four-letter word: Victimization and romantic relationships in adolescence. *Child Maltreatment, 5*, 293-298.

Follingstad, D. R. (2007). Rethinking current approaches to psychological abuse: Conceptual and methodological issues. *Aggression and Violent Behavior, 12*, 439-458.

Follingstad, D. R. (2011). A measure of severe psychological abuse normed on a nationally representative sample of adults. *Journal of Interpersonal Violence, 26*, 1194-1214.

Foshee, V. A. (1996). Gender differences in adolescent dating abuse prevalence, types and injuries. *Health Education Research, 11*, 275-286.

Foshee, V. A., Bauman, K. E., Linder, F., Rice, J., & Wilcher, R. (2007). Typologies of adolescent dating violence: Identifying typologies of adolescent dating violence perpetration. *Journal of Interpersonal Violence, 22*, 498-519.

Foshee, V. A., Benefield, T., Suchindran, C., Ennett, S. T., Bauman, K. E., Karriker-Jaffe, K. J., … Mathias, J. (2009). The development of four types of adolescent dating abuse and selected demographic correlates. *Journal of Research on Adolescence, 19*, 380-400.

Foshee, V. A., & Reyes, H. L. M. (2007). Primary prevention of adolescent dating abuse perpetration: When to begin, whom to target, and how to do it. In D. J. Whitaker & J. R. Lutzker (Eds.), *Preventing partner violence: Research and evidence-based intervention strategies* (pp. 141-168). Washington, DC: American Psychological Association.

Foshee, V. A., & Reyes, H. L. M. (2011). Dating abuse: Prevalence, consequences, and predictors. In R. J. R. Levesque (Ed.), *Encyclopedia of Adolescence* (pp. 602-615). New York, NY: Springer Science + Business Media.

Furman, W., Low, S., & Ho, M. J. (2009). Romantic experiences and psychosocial adjustment in middle adolescence. *Journal of Clinical Child & Adolescent Psychology, 38*, 75-90.

Furman, W., & Shaffer, L. (2006). The role of romantic relationships in adolescent development. In A. C. Crouter & A. Booth (Eds.), *Romance and sex in adolescence and emerging adulthood: Risks and opportunities* (pp. 3-22). Mahwah, NJ: Lawrence Erlbaum Associates.

Furman, W., & Shomaker, L. B. (2008). Patterns of interaction in adolescent romantic relationships: Distinct features and links to other close relationships. *Journal of Adolescence, 31*, 771-788.

Giordano, P. C., Manning, W. D., & Longmore, M. A. (2006). Adolescent romantic relationships: An emerging portrait of their nature and developmental significance. In A. C. Crouter & A.

Booth (Eds.), *Romance and sex in adolescence and emerging adulthood: Risks and opportunities* (pp. 127-150). Mahwah, NJ: Lawrence Erlbaum Associates.

Giordano, P. C., Soto, D. A., Manning, W. D., & Longmore, M. A. (2010). The characteristics of romantic relationships associated with teen dating violence. *Social Science Research, 39*, 863-874.

Gómez, A. M. (2011). Testing the cycle of violence hypothesis: Child abuse and adolescent dating violence as predictors of intimate partner violence in young adulthood. *Youth & Society, 43*, 171-192.

Gray, H. M., & Foshee, V. (1996). Adolescent dating violence: Differences between one-sided and mutually violent profiles. *Journal of Interpersonal Violence, 12*, 126-141.

Joyner, K., & Campa, M. (2006). How do adolescent relationships influence the quality of romantic and sexual relationships in young adulthood? In A. C. Crouter & A. Booth (Eds.), *Romance and sex in adolescence and emerging adulthood: Risks and opportunities* (pp. 93-102). Mahwah, NJ: Lawrence Erlbaum Associates.

Knoble, N. B., Capaldi, D. M., Shortt, J. W., & Kim, H. K. (2012). *Risk factors for adolescent dating violence: A systematic review.* Poster presented at the 14th Society for Research on Adolescence Biennial Meeting, Vancouver, BC.

Meier, A., & Allen, G. (2009). Romantic relationships from adolescence to young adulthood: Evidence from the National Longitudinal Study of Adolescent Health. *The Sociological Quarterly, 50*, 308-335.

Miller, S., Williams, J., Cutbush, S., Gibbs, D., Clinton-Sherrod, M., & Jones, S. (2013). Dating violence, bullying, and sexual harassment: Longitudinal profiles and transitions over time. *Journal of Youth and Adolescence, 42*, 607-618.

Mischo, C. (2005). Promoting perspective coordination by dilemma discussion: The effectiveness

of classroom group discussion on interpersonal negotiation strategies of 12-year-old students. *Social Psychology of Education, 8*, 41-63.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674-701.

Neemann, J., Hubbard, J., & Masten, A. S. (1995). The changing importance of romantic relationship involvement to competence from late childhood to late adolescence. *Development and Psychopathology, 7*, 727-750.

Offenhauer, P., & Buchalter, A. (2011). *Teen dating violence: A literature review and annotated bibliography* (DOJ Document No. 235368). Washington, DC: Library of Congress.

Orpinas, P., Hsieh, H., Song, X., Holland, K., & Nahapetyan, L. (2013). Trajectories of physical dating violence from middle to high school: Association with relationship quality and acceptability of aggression. *Journal of Youth and Adolescence, 42*, 551-565.

Pavao, J., Alvarez, J., Baumrind, N., Induni, M., & Kimerling, R. (2007). Intimate partner violence and housing instability. *American Journal of Preventive Medicine, 32*, 143-146.

Roberts, T. A., Klein, J. D., & Fisher, S. (2003). Longitudinal effect of intimate partner abuse on high-risk behavior among adolescents. *Archives of Pediatrics and Adolescent Medicine, 157,* 880-881.

Roisman, G. I., Masten, A. S., Coatsworth, J. D., & Tellegen, A. (2004). Salient and emerging developmental tasks in the transition to adulthood. *Child Development, 75*, 123-133.

Seiffge-Krenke, I., & Burk, W. J. (2013). Friends or lovers? Person- and variable-oriented perspectives on dyadic similarity in adolescent romantic relationships. *Journal of Social and Personal Relationships, 30*, 711-733.

Selman, R. L. (1980). *The growth of interpersonal understanding: Developmental and clinical analyses.* New York: Academic Press.

Sherer, M. (2009). The nature and correlates of dating violence among Jewish and Arab youths in Israel. *Journal of Family Violence, 24,* 11-26.

Smith, P. H., White, J. W., & Holland, L. J. (2003). A longitudinal perspective on dating violence among adolescent and college-age women. *American Journal of Public Health, 93*, 1104-1109.

Staggs, S. L., & Riger, S. (2005). Effects of intimate partner violence on low-income women's health and employment. *American Journal of Community Psychology, 36*, 133-145.

Steinberg, L. (2008a). *Adolescence*. New York, NY: McGraw-Hill.

Steinberg, L. (2008b). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28*, 78-106.

Sticca, F., & Perren, S. (2013). Is cyberbullying worse than traditional bullying? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. *Journal of Youth and Adolescence, 42*, 739-750.

Straus, M.A., Hamby, S.L., Boney-McCoy, S., & Sugarman, D.B. (1996). The revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283-316.

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. New York: W.W. Norton & Company.

Vagi, K. J., Rothman, E. F., Latzman, N. E., Tharp, A. T., Hall, D. M., & Breiding, M. J. (2013). Beyond correlates: A review of risk and protective factors for adolescent dating violence perpetration. *Journal of Youth and Adolescence, 42*, 633-649.

Williams, T. S., Connolly, J., Pepler, D., Craig, W., & Laporte, L. (2008). Risk models of dating aggression across different adolescent relationships: A developmental psychopathology approach. *Journal of Consulting and Clinical Psychology, 76*, 622-632.

Wolfe, D. A., Scott, K., Reitzel-Jaffe, D., Wekerle, C., Grasley, C., & Straatman, A. (2001).

Development and validation of the Conflict in Adolescent Dating Relationships Inventory. *Psychological Assessment, 13*, 277-293.

Wolitzky-Taylor, K. B., Ruggiero, K. J., Danielson, C. K., Resnick, H. S., Hanson, R. F., Smith, D. W., . . . Kilpatrick, D. G. (2008). Prevalence and correlates of dating violence in a national sample of adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*, 755-762.

Zimmer-Gembeck, Siebenbruner, J., & Collins, W. A. (2001). Diverse aspects of dating: Associations with psychosocial functioning from early to middle adolescence. *Journal of Adolescence, 24*, 313-336.

CHAPTER TWO

RE-VICTIMIZATION AFTER TEEN DATING VIOLENCE

Re-victimization after Teen Dating Violence

Deinera Exner-Cortens

John Eckenrode

John Bunge

Emily Rothman

Author Note

Deinera Exner-Cortens, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

John Eckenrode, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

John Bunge, Department of Statistical Science, Cornell University.

Emily Rothman, Department of Community Health Sciences, Boston University School of Public Health.

This research was supported in part by a Doctoral Foreign Study Award from the Canadian Institutes of Health Research. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Correspondence concerning this article should be addressed to Deinera Exner-Cortens, CAMH Centre for Prevention Science, 100 Collip Circle, Suite 100, London, ON N6G-4X8. Contact: dme56@cornell.edu

Abstract

Violence experienced in adolescent dating relationships is an important public health issue, and emerging research demonstrates that individuals who experience victimization in adolescence are more likely to be re-victimized in future relationships: this is especially concerning given the association of adult intimate partner violence with adverse health, economic and family outcomes. However, past work on this topic is limited by potential confounding, and lack of assessment of potential mediators of this relationship. Thus, the present study used data from the National Longitudinal Study of Adolescent Health to explore pathways to re-victimization, adjusting for confounding using a high-dimension propensity score. Results indicate that dating violence experienced during adolescence is indirectly associated with intimate partner violence experienced 12 years later, through the experience of intimate partner violence at 5.5 year follow-up. This finding suggests that violence experienced in intimate relationships may become chronic, and future work should focus on understanding and addressing this pattern.

Re-victimization after Teen Dating Violence

Evidence that teen dating violence (TDV)—the experience of physical, psychological and/or sexual aggression in adolescent dating relationships—is a substantial public health problem continues to grow. National prevalence of physical dating violence has remained constant over the last decade (~10%; CDC, 2012), and several longitudinal studies have established the association of TDV with future adverse outcomes (Ackard, Eisenberg, & Neumark-Sztainer, 2007; Brown, Cosgrave, Killackey, Purcell, Buckby, & Yung, 2009; Exner-Cortens, Eckenrode, & Rothman, 2013; Roberts, Klein, & Fisher, 2003; Teitelman, Ratcliffe, Dichter, & Sullivan, 2008; van Dulmen et al., 2012). Recently, Exner-Cortens et al. (2013) demonstrated that approximately five years post-victimization, males who had experienced TDV reported more antisocial behaviors, suicidal ideation, and marijuana use than their non-victimized peers, and females who had experienced TDV reported more heavy episodic drinking, depressive symptomatology, suicidal ideation and smoking. Further, several studies document the association of TDV with re-victimization in adolescence and young adulthood (Cui, Gordon, Ueno, & Fincham, 2013; Exner-Cortens et al., 2013; Gómez, 2011; Halpern, Spriggs, Martin, & Kupper, 2009; Rich, Gidycz, Warkentin, Loh, & Weiland, 2005; Smith, White, & Holland, 2003; Spriggs, Halpern, & Martin, 2009a; Teitelman et al., 2011; Williams, Connolly, Pepler, Craig, & Laporte, 2008). For example, in a sample of 621 Canadian adolescents (mean age=15.35), 13% of participants experienced dating violence in two different relationships over a one-year period (Williams et al., 2008), while using National Longitudinal Study of Adolescent Health (Add Health) data, Spriggs et al. (2009a) found that 32% of those reporting TDV victimization were also recipients of intimate partner violence (IPV) victimization in young adulthood. Given the broad consequences of IPV victimization on health and well-being (Breiding, Black, & Ryan, 2008; Coker et al., 2002), healthcare utilization (Rivara et al., 2007), economic outcomes

(NCIPC, 2003; Max, Rice, Finkelstein, Bardwell, & Leadbetter, 2004), and children in the household (Dube, Anda, Felitti, Edwards, & Williamson, 2002; Kitzmann, Gaylord, Holt, & Kenny, 2003), the potential for re-victimization following TDV is especially concerning.

While prior studies have demonstrated a link between TDV and future IPV victimization, none have investigated possible mechanisms that might explain risk for re-victimization; understanding pathways to re-victimization is important to improving secondary prevention programs and treatment. It is plausible that the experience of re-victimization itself (e.g., victimization occurring in both adolescence and young adulthood) may place someone at risk for chronic victimization (i.e., experiencing victimization in multiple relationships over time). However, other work has demonstrated associations between TDV and future depressive symptomatology and substance use (Ackard et al., 2007; Exner-Cortens et al., 2013; Roberts et al., 2003), as well as between depressive symptomatology, substance use and future IPV victimization (e.g., Lehrer, Buka, Gortmaker, & Shrier, 2006; Testa, Livingston, & Leonard, 2003), and so it is also possible that depression and/or substance use may mediate the TDV – IPV link.

Beyond potential pathways that may explain risk for re-victimization, past work is also limited by the potential confounding of the TDV – IPV relationship; these studies typically only controlled for socio-demographics (e.g., race/ethnicity, SES), although research has demonstrated that risk for TDV exists across multiple levels of the social ecology (e.g., Knoble, Capaldi, Shortt, & Kim, 2012; Lewis & Fremouw, 2001; Offenhauer & Buchalter, 2011; Sherer, 2009). Thus, it appears that teens who experience dating violence differ from their non-victimized peers on multiple risk indicators, and that controlling for socio-demographics alone will not necessarily ensure ignorable treatment assignment. Propensity score matching, where the goal is to balance the distribution of covariates in the treatment (i.e., victims of TDV) and control

(i.e., persons who have not experienced any TDV) groups, is one alternative that would help address this limitation, since the propensity score method allows for adjustment on a large number of covariates, with the purpose of meeting the assumption of strongly ignorable treatment assignment: the propensity score defines the distance between two individuals on selected covariates by summarizing the covariates into one quantity, the probability of being treated (Stuart, 2010). Although Cui et al. (2013) did use propensity scores in their investigation of re-victimization after dating violence, they included primarily socio-demographic variables in their propensity score model.

Thus, the present study investigated whether the experience of psychological and physical TDV victimization in adolescence was associated with physical IPV victimization approximately 12 years later, using a comprehensive propensity score for adjustment and matching. We also investigated if the relationship between TDV and adult IPV victimization was mediated by victimization, depression and/or substance use that occurred approximately 5.5 years after TDV was assessed.

Method

Data

This study analyzed data from the Add Health dataset. Add Health was designed to study determinants of health and risk behaviors in a nationally representative sample of U.S. adolescents. In 1994, participants were selected from 80 high schools and 52 middle schools, stratified with respect to region of country, urbanicity, school size, school type and ethnicity. At Wave 1 (1994-1995), adolescents in grades 7-12 participated in a structured in-home interview. Adolescents were re-interviewed in 1996 (Wave 2), 2001-2002 (Wave 3) and 2007-2008 (Wave 4). Average time between the Wave 1 and 2 interviews was approximately 11 months, between Wave 2 and Wave 3, 5.5 years, between Wave 3 and Wave 4, 6.5 years, and between Wave 2

and Wave 4, 12 years. Data from the Wave 1 School Administrator Questionnaire, the Wave 1 Adolescent In-School Questionnaire, and the Wave 1 Parent In-Home Questionnaire were also used in this paper.

Sample

The analytic sample was restricted to adolescents who participated in the in-home interviews at Waves 1, 2, 3 and 4 (*n*=9,421). Participants were included if they reported that they 1) had been in a heterosexual dating or sexual relationship between the Wave 1 and 2 interviews, and weren't married at either Wave 1 or Wave 2 (*n*=6,210; Halpern et al., 2009; Spriggs et al., 2009a); 2) were 18 years or younger at Wave 2 (*n*=5,746); 3) had answered Wave 1 and Wave 2 audio computer assisted self-interview (A-CASI) questions honestly (*n*=5,379; Roberts & Klein, 2003); and 4) has experienced their first reported victimization incident after the Wave 1 interview, or had never been victimized by Wave 2 (*n*=4,582). For propensity score matching, included participants also needed to have complete data on all covariates (*n*=3,961). Complete case analysis resulted in the exclusion of 13.6% of the eligible sample.

Measures

Teen dating violence victimization. At Wave 2, participants identified up to 3 romantic and 3 sexual relationships occurring since the Wave 1 interview. Participants were asked about violence victimization experienced in each relationship using A-CASI (other data collected by A-CASI are indicated in Appendix 2.A). Dating violence was measured using 5 items from the revised Conflict Tactics Scale (CTS2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996). Participants were asked if a partner had ever 1) called them names, insulted them or treated them disrespectfully in front of others; 2) sworn at them; 3) threatened them with violence; 4) pushed or shoved them; or 5) thrown something at them that could hurt. For the present analyses, a dichotomous variable was created, indicating whether participants endorsed the particular

victimization item in any of their romantic or sexual relationships. Because propensity score analysis requires that all covariates be collected pre-treatment (Stuart, 2010), individuals included in the present analyses needed to have experienced their first victimization incident after the Wave 1 interview date (*n*=737).

While we planned to examine associations with re-victimization in the group experiencing psychological violence only (item(s) 1, 2 and/or 3) and the group experiencing physical and psychological violence (item(s) 1, 2 and/or 3 *and* item(s) 4 and/or 5), as we did in our prior work with these data (Exner-Cortens et al., 2013), these groups were too small to consider separately, due to the sample size reduction created by only including individuals who experienced victimization after the Wave 1 interview. Instead, we explored all associations in the group experiencing any dating violence victimization (i.e., individuals endorsing any of items 1 – 5). The comparison group was adolescents reporting having dating or sexual partners but no dating violence at Wave 2.

Pre-treatment covariates. Appendix 2.A lists the 53 variables that were considered for the propensity score model. The list was generated by examining past empirical studies of predictors of TDV; longitudinal studies were the primary source for this list. We also consulted a recent systematic review of TDV predictors (Knoble, Capaldi, Shortt, & Kim, 2012) and considered theories of dating violence (e.g., Exner-Cortens, 2013). Empirical studies demonstrating an association between the predictor variable and TDV are cited in the first column of Appendix 2.A (this list is not intended to be comprehensive, but rather to provide examples of empirical support for associations). To determine if predictors were also theoretically or empirically associated with IPV victimization, and should therefore be included on our list of propensity score variables, we consulted a recent systematic review of predictors of IPV (Capaldi, Knoble, Shortt, & Kim, 2012), as well as considered theories of domestic violence

(e.g., social learning theory). In order to ensure that included variables represented the multiple contexts that might influence risk for dating and intimate partner violence, variables were organized according to the social-ecological model, as described by Dahlberg and Krug (2002).

Wave 3 mediators.

Depression. Nine items from the 20-item Centers for Epidemiological Studies—Depression Scale (CES-D) were used to assess depressive symtomatology (Radloff, 1977). Participants were asked if they had experienced particular feelings (e.g., you felt depressed) in the past 7 days. The nine items included in Add Health represent the four subscales of the CES-D (Crockett, Randall, Shen, Russell, & Driscoll, 2005). Items were summed; higher scores indicate greater depressive symptomatology (M(SE) = 4.29 (0.10), range, 0-25; $\alpha = 0.81$). In addition to the sum score, a cut score was also considered, in order to more clearly define a threshold for depression. Using the truncated 9-item scale, Primack et al. (2009) applied cut-offs of ≥11 for females and ≥10 for males; these cut-offs were also used here. In this sample, 8.9% (*n*=360) of participants exceeded the cut-off.

Substance use. Participants reported on smoking behavior in the past 30 days. This variable was dichotomized, indicating smoking on one or more days. To assess drinking behavior, participants reported how many times they had drank five or more drinks in a row in the past year. Heavy episodic drinking was defined as having at least two to three such episodes a month for each of the preceding 12 months (1=*Yes,* 0=*No*). Illicit substance use was divided into two categories: marijuana use and other drug use (cocaine, crystal meth, other illegal drugs (e.g., LSD, ecstasy), injection drugs) in the past 12 months. Both variables were dichotomized, indicating any marijuana or other drug use.

Intimate partner violence victimization. Participants reported on physical violence victimization occurring in romantic and sexual relationships in the past 12 months. Physical IPV

items were derived from the CTS2 (Straus et al., 1996); participants were asked if a partner had 1) threatened them with violence, pushed or shoved them, or thrown something at them that could hurt or 2) slapped, hit or kicked them. Responses were measured on a 7-point Likert-type scale, ranging from never to more than 20 times. A dichotomous variable indicates whether participants endorsed either Wave 3 IPV item.

Wave 4 outcome variable.

Intimate partner violence victimization. At Wave 4, participants were asked to report on past year physical violence victimization occurring in their current or most recent romantic or sexual relationship. Participants were skipped out of this section of the questionnaire if they did not have a partner, if none of their partners satisfied the criteria for inclusion, or if they refused to answer (n=485). IPV items were derived from the CTS2 (Straus et al., 1996), and were the same as the items used at Wave 3. Responses were measured on a 7-point Likert-type scale, ranging from never to more than 20 times. A dichotomous variable indicates the report of any physical IPV at Wave 4. While we also created chronicity scores (i.e., scores that represent the number of times the individual was victimized, instead of a dichotomous indicator; Straus, 2004), due to the severe positive skew in this variable, the dichotomous prevalence score was used in all analyses.

Analysis

Propensity score matching was used to create matched treatment and control groups (where the experience of any TDV was considered the treatment variable). To determine which variables to include in the propensity score model, a directed acyclic graph was created, and vertices that were connected to both Wave 2 TDV and Wave 4 IPV by an edge were included in the initial model. Because propensity score analysis requires complete data, and because multiple imputation was not an option in this dataset (Exner-Cortens et al., 2013), we also excluded variables with large amounts of missing data from the initial model.

Pre-treatment covariates included in the propensity score model are indicated in Appendix 2.A (n=23). In order to determine if variable exclusion decisions were appropriate, pre- and post-match balance was checked for all excluded variables; if any of these variables had demonstrated imbalance post-match (see below), we would have re-specified the model to include them. However, checks on excluded variables demonstrated post-match balance, and so this was not necessary. Prior to propensity score matching, we also graphically assessed the region of common support for our treatment and control groups. This graph indicated a substantial region of overlap between the pre-match propensity score distributions of the treatment and control group.

Because Add Health is a complex survey design, we generated the propensity scores using a logistic regression model that accounted for design effects (i.e., post-stratification, unequal probability of selection, and clustering; Chantala & Tabor, 1999), using R's Survey package (The R Foundation for Statistical Computing. Available at www.r-project.org, 2010). To created a matched group using these scores, we used the R package MatchIt (Ho, Imai, King, & Stuart, 2011). Good balance was found for a 1:2 nearest neighbor match without replacement on the propensity score (caliper=0.20), an exact match on gender, and a Mahalanobis match on child maltreatment, family structure and race (Table 2.1). Matching was done using the linear propensity score as the distance measure (i.e., $|\text{logit}(e_k) - \text{logit}(e_j)|$, where e_k and e_j are the propensity scores for person k and j, respectively; Stuart, 2010). Covariate balance was assessed using the standardized difference in the means; while Stuart (2010) suggests that an absolute standardized difference < 0.25 is needed for regression adjustment to be trustworthy (p. 11), we used a more stringent level of 0.10 for model-included covariates, per Hughes, Chen, Thoemmes and Kwok (2010). Using the standardized difference, balance was checked for all model-included covariates (including all interaction terms), as well as on all model-excluded covariates

and interactions with a pre-match standardized difference > |0.25|. Balance was also assessed using graphs (QQ plots of each covariate, and histograms and jitter plots of the linear propensity score), and by assessing the ratio of variances of the propensity score in the treatment and control groups (Table 2.1). As discussed by Rubin (2001), this ratio should be close to one. The final sample size in the matched sample was 2,161 (59.1% female, *n*=732 treatment units); five treatment units were discarded during the match. Post-match graphical analysis of the common support region indicated that all control individuals had been selected from within this region.

Table 2.1
Summary of Balance Checks

	Pre-match (*n*=3,961)	**Post-match (*n*=2,161)**
Standardized difference, % (n)		
$x < \|.01\|$	4.6 (35)	21.3 (163)
$\|.01\| < x < \|.05\|$	18.8 (144)	61.7 (473)
$\|.05\| < x < \|.10\|$	24.1 (185)	16.0 (123)
$x > \|.10\|$	52.5 (403)	1.0 (8)
Variance		
Variance ratio	0.75	0.90
Variance, treatment	0.43	0.46
Variance, control	0.39	0.34

x=standardized difference. Standardized difference ranges presented in this table are for all model-included covariates and their interaction terms (*n*=767 variables).

Since propensity score matching by definition violates the survey design (i.e., units are discarded from the sample, resulting in a loss of information required to accurately compute the variance; Chantala, 2006), we conducted all outcome analyses using two samples: 1) the full, pre-match population (*n*=3,961), accounting for survey design effects, and controlling for the propensity score, race, family structure and child maltreatment; and 2) the matched group (*n*=2,161), not accounting for design effects, and controlling for the propensity score only (Stuart, 2010). Outcome analyses for the matched sample were weighted to account for the 1:2 nearest neighbor match.

All outcome analyses were performed in Mplus v.6.12, using path models (Figure 2.1).

Models were estimated under WLSMV with a probit link function; this model treats observed binary endogenous variables as continuous latent response variable, such that,

$$y_i^* = \begin{cases} 0 \; if \; y_i^* \leq \tau \\ 1 \; if \; y_i^* > \tau \end{cases}$$

where τ is a threshold that defines whether or not the event was observed (Múthen, 2011). Thus, parameters for all endogenous variables are interpreted as linear, representing an increase or decrease in the latent probability of the event. Mediation was assessed using the MODEL INDIRECT command in Mplus, where for the model,

$$y_i^* = \beta_0 + \beta_1 m_i^* + \beta_2 x_i + \beta_3 c_i + \varepsilon_{1i},$$

$$m_i^* = \gamma_1 x_i + \gamma_2 c_i + \varepsilon_{2i}$$

the natural direct effect is β_2 and the natural indirect effect is $\beta_1\gamma_1$, conditional on the covariate *c* (Múthen, 2011). Bias-corrected bootstrap confidence intervals were generated for all parameter estimates (Fritz & MacKinnon, 2007).

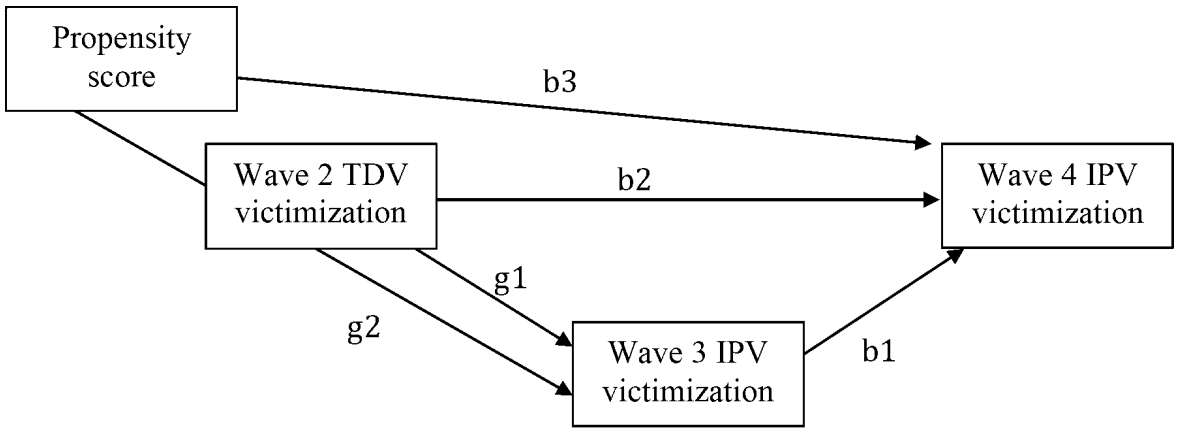

Figure 2.1. Simplified path diagram of associations tested in Tables 2.5 and 2.6.

To explore the impact of missing data, individuals with any missing data on covariates were compared to individuals with no missing data. There was no difference between missing and non-missing individuals on age, puberty, child maltreatment or dating violence victimization. However, missing individuals did differ on sex (more likely to be male), race

(more likely to be Hispanic or Black, non-Hispanic), family structure (more likely to be in a household without biological or step-parents), and parental education (more likely to be less than high school).

Bivariate associations between TDV victimization and other variables were explored using t-tests or χ^2 tests, as appropriate. All analyses (bivariate and path analysis) were stratified by sex, and all results were evaluated at $p < .05$. This study was reviewed by the Cornell University Institutional Review Board and deemed exempt.

Results

Sample Characteristics

Sample demographics for the pre- and post-match samples are presented in Table 2.2. Prior to the match, treatment and control group participants differed on age, family structure, child maltreatment history, pubertal status, and the overall propensity score. After the match, there were no significant differences on any of these variables (Table 2.2). In the pre-match sample, prevalence of dating violence occurring in the period between the Wave 1 and 2 interviews (~11 months) was 18.6%.

Table 2.2
Sociodemographics by Group

	% (n)[a]			
	Pre-match (*n*=3,961)		**Post-match (*n*=2,161)**	
	Treatment (n=737)	*Control (n=3,224)*	*Treatment (n=732)*	*Control (n=1,429)*
Propensity score, mean (SE)	-1.26 (0.04)	-1.68 (0.02)***	-1.32 (0.02)	-1.33 (0.02)
Wave 1 age, mean (SE)	15.70 (0.11)	15.42 (0.10)**	15.86 (0.05)	15.92 (0.04)
Sex				
Male	43.7 (302)	46.8 (1315)	40.7 (298)	40.7 (585)
Female	56.3 (435)	53.2 (1909)	59.3 (434)	59.3 (844)

(Table 2.2 continues)

(Table 2.2 continued)

	% (n)[a]			
	Pre-match (*n*=3,961)		**Post-match (*n*=2,161)**	
	Treatment (n=737)	*Control (n=3,224)*	*Treatment (n=732)*	*Control (n=1,429)*
Race/ethnicity				
White, non-Hispanic	67.4 (421)	71.7 (1897)	57.0 (417)	56.1 (804)
Black, non-Hispanic	14.2 (146)	12.1 (568)	19.9 (146)	19.7 (278)
Hispanic	10.5 (101)	10.0 (468)	13.7 (100)	14.5 (207)
Other	8.0 (69)	6.2 (291)	9.4 (69)	9.8 (140)
Family structure				
Two biological parents	54.1 (393)	59.1 (1864)**	53.1 (389)	54.6 (783)
Step-family	15.8 (122)	17.5 (575)	16.5 (121)	16.7 (238)
Single parent	25.5 (193)	21.0 (703)	26.4 (193)	25.6 (363)
Other	4.6 (29)	2.3 (82)	4.0 (29)	3.1 (45)
Parental education (SES)				
Less than high school	11.1 (83)	10.0 (339)	11.2 (82)	11.2 (157)
High school graduate	30.1 (223)	29.8 (894)	30.5 (223)	27.2 (389)
Some post-secondary	21.3 (151)	22.5 (707)	20.4 (149)	23.0 (333)
College graduate	25.5 (189)	25.0 (823)	25.7 (188)	25.1 (356)
Post-college	12.1 (91)	12.7 (461)	12.3 (90)	13.4 (194)
Child maltreatment				
Yes	24.3 (185)	19.0 (621)**	24.9 (182)	23.0 (326)
No	75.7 (552)	81.0 (2603)	75.1 (550)	77.0 (1103)
Pubertal status				
> 1 SD above mean	17.0 (121)	13.9 (465)**	16.0 (117)	17.3 (242)
Within ± 1 SD of mean	73.2 (550)	72.5 (2326)	75.3 (549)	73.8 (1041)
< 1 SD below mean	9.8 (63)	13.6 (388)	8.6 (63)	8.9 (127)
Type of dating violence				
Psychological only	12.2 (500)	n/a	23.1 (499)	n/a
Physical only	1.3 (49)	n/a	2.2 (48)	n/a
Physical and psychological	5.1 (188)	n/a	8.6 (185)	n/a

*$p < .05$; **$p < .01$; ***$p < .001$

[a]Unless otherwise noted. For the pre- and post-match samples, percentages and means are weighted, number of subjects is unweighted. However, weights were different in these two samples (in the pre-match sample, the weight

reflects unequal probability of selection, while in the post-match sample, the weight is for the control units, and is proportional to the number of treatment units to which the control unit was matched). Because of this, percentages and means are not directly comparable between the pre- and post-match samples.

Bivariate Associations

Bivariate associations between the treatment (Wave 2 dating violence victimization), potential Wave 3 mediators (substance use, depression and IPV victimization) and Wave 4 outcome (IPV victimization) are presented in Tables 2.3 and 2.4. For all presented analyses, depression was indicated using the cut score and not the continuous variable. In Tables 2.3 and 2.4, both the weighted and unweighted percentages are presented for the full (pre-match) sample, in order to facilitate comparison with the post-match sample (percentages between the unweighted full sample and unweighted matched sample are comparable).

In the full, survey-weighted female sample, Wave 2 dating violence victimization was significantly associated with Wave 3 smoking, marijuana use, other drug use, depression and IPV victimization, but was not associated with Wave 4 IPV victimization (Table 2.3). In the matched female sample, all associations were reduced to non-significance, except the association between Wave 2 dating violence victimization and Wave 3 IPV victimization (Table 2.3).

Table 2.3
Bivariate Associations: Female Participants

	% yes (n)					
	Full sample, survey-weighted		**Full sample, unweighted (frequencies only)**		**Matched sample, unweighted**	
	Any TDV (*n*=435)	No TDV (*n*=1909)	Any TDV (*n*=435)	No TDV (*n*=1909)	Any TDV (*n*=434)	No TDV (*n*=844)
Wave 3 Mediators						
Smoking	44.1 (174)	35.8 (587)*	40.3 (174)	30.9 (587)	40.1 (173)	35.6 (298)
Heavy episodic drinking	21.6 (83)	18.2 (279)	19.4 (83)	14.9 (279)	19.2 (82)	15.8 (130)
Marijuana use	39.1 (156)	31.2 (524)*	36.4 (156)	27.7 (524)	36.3 (155)	31.0 (257)
Other drug use	16.4 (60)	11.1 (180)*	14.0 (60)	9.5 (180)	13.8 (59)	11.3 (94)
Depression	16.0 (58)	8.9 (185)**	13.4 (58)	9.7 (185)	13.2 (57)	11.9 (100)
Intimate partner violence victimization	44.2 (153)	26.3 (450)***	39.7 (153)	26.8 (450)	39.8 (153)	27.3 (205)***
Wave 4 Outcome Variable						
Intimate partner violence victimization	14.0 (64)	11.5 (227)	15.1 (64)	12.3 (227)	15.2 (64)	13.6 (110)

^$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$

Note: In the full, survey-weighted sample (Column 1), percentages are weighted, and number of subjects is unweighted. In order to provide comparable full (i.e., pre-match) sample and matched sample frequencies, full sample data were also summarized using unweighted percentages (Column 2): Column 2 frequency data are provided for comparison purposes only, and so significant associations in the unweighted full data are not presented. Percentages and number of subjects in the matched sample (Column 3) are also unweighted (weighted percentages were calculated for the matched sample (data not shown), and were very similar to the unweighted percentages, with no difference in significance of associations).

In the full, survey-weighted male sample, Wave 2 dating violence victimization was significantly associated with Wave 3 marijuana use, other drug use and IPV victimization, but was not associated with Wave 4 IPV victimization (Table 2.4). In the matched sample, all associations were reduced to non-significance, except the association between Wave 2 dating violence victimization and Wave 3 IPV victimization (Table 2.4).

Table 2.4
Bivariate Associations: Male Participants

	% yes (n)					
	Full sample, survey-weighted		**Full sample, unweighted (frequencies only)**		**Matched sample, unweighted**	
	Any TDV (*n*=302)	No TDV (*n*=1315)	Any TDV (*n*=302)	No TDV (*n*=1315)	Any TDV (*n*=298)	No TDV (*n*=585)
Wave 3 Mediators						
Smoking	47.6 (137)	40.6 (480)^	45.8 (137)	37.0 (480)	45.1 (133)	39.7 (228)
Heavy episodic drinking	42.6 (112)	37.7 (462)	37.6 (112)	35.9 (462)	37.1 (109)	37.1 (211)
Marijuana use	49.8 (144)	41.4 (505)*	48.2 (144)	38.8 (505)	47.8 (141)	43.6 (251)
Other drug use	28.3 (66)	17.5 (208)***	22.1 (66)	16.0 (208)	21.4 (63)	19.5 (112)
Depression	8.7 (30)	6.9 (87)	10.0 (30)	6.6 (87)	9.5 (28)	8.4 (49)
Intimate partner violence victimization	37.5 (97)	20.0 (205)***	37.7 (97)	18.4 (205)	37.4 (95)	19.0 (94)***
Wave 4 Outcome Variable						
Intimate partner violence victimization	25.5 (72)	20.6 (230)	25.0 (72)	18.2 (230)	24.9 (71)	20.6 (115)

^$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$
Note: In the full, survey-weighted sample (Column 1), percentages are weighted, and number of subjects is unweighted. In order to provide comparable full (i.e., pre-match) sample and matched sample frequencies, full sample data were also summarized using unweighted percentages (Column 2): Column 2 frequency data are provided for comparison purposes only, and so significant associations in the unweighted full data are not presented. Percentages and number of subjects in the matched sample (Column 3) are also unweighted (weighted percentages were calculated for the matched sample (data not shown), and were very similar to the unweighted percentages, with no difference in significance of associations).

While we had originally intended to assess mediation using substance use and depression, small cell sizes for some of these variables, as well as the lack of any significant association between Wave 2 dating violence and substance use or depression in the matched sample, led to the decision to examine mediation by Wave 3 IPV victimization only.

Path Analysis

For both males and females, dating violence victimization at Wave 2 was significantly associated with IPV at Wave 3, in both the full (survey-weighted) and matched samples (Tables 2.5 and 2.6).[2] In the survey-weighted sample, the predicted probability of Wave 3 IPV given a history of dating violence victimization in females was 0.44 (range, 0.29-0.66), while the predicted probability for non-victims was 0.26 (range, 0.14-0.53). For males, the predicted probability of Wave 3 IPV given a history of dating violence victimization was 0.37 (range, 0.17-0.75), while the predicted probability for non-victims was 0.20 (range, 0.082-0.59).

Dating violence at Wave 2 was not directly predictive of Wave 4 IPV victimization in either the survey-weighted or matched samples (Tables 2.5 and 2.6). However, IPV at Wave 3 was significantly associated with IPV at Wave 4 for both females (Table 2.5) and males (Table 2.6), and was part of a significant indirect effect (where the indirect effect represents the amount by which the outcome, y^*, is expected to increase or decrease indirectly through the mediator, m^*, per a unit change in the independent variable, x; Preacher & Kelley, 2011, p. 99). Thus, for both males and females who experienced dating violence at Wave 2 (x=1), the probability of experiencing Wave 4 IPV victimization increased indirectly through the experience of Wave 3 IPV victimization. The distribution of the predicted probabilities for experiencing Wave 4 IPV in the treatment and control group are presented in Figure 2.2 (females) and Figure 2.3 (males),

[2] All path models were checked for treatment-mediator interactions using nested log-likelihood tests under MLR estimation with a probit link (Múthen, 2011). None of these interactions were significant, nor did they improve the fit, and so they were removed from the model.

using parameter estimates generated in the full (survey-weighted) sample.[3]

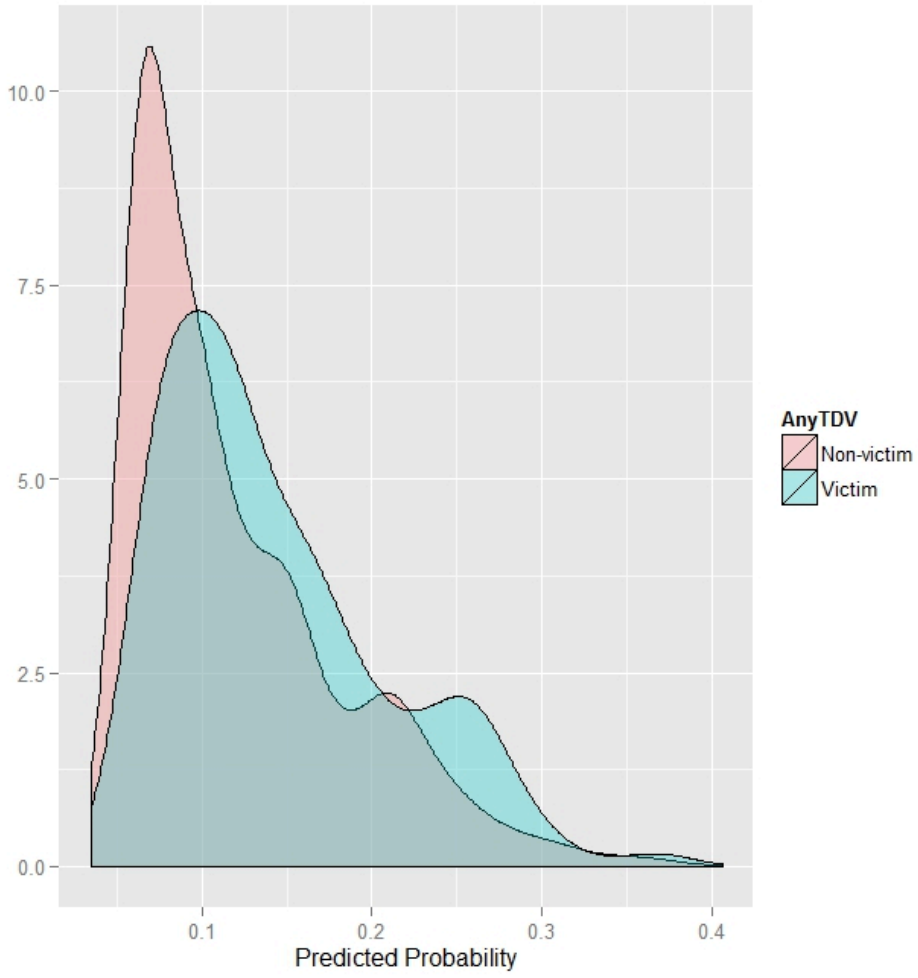

Figure 2.2. Predicted probability distribution for female participants of the probability of Wave 4 IPV victimization. Predicted probabilities were created using parameter estimates from the survey-weighted model for the full sample.

[3] Using coefficients standardized with respect to the latent unobserved variable *y** (such that the indirect effect coefficient represents the change in *y** in *y* standard deviation units when *x* changes from zero to one (Múthen & Múthen, 2012, p. 722; data not shown), the indirect effect for girls was similar in magnitude whether estimated using survey weighting (b=0.09) or in the matched group (b=0.08). The indirect coefficients for boys were larger in magnitude than the coefficients for girls, and were slightly different when calculated under survey weighting (b=0.13) vs. using the matched group (b=0.19).

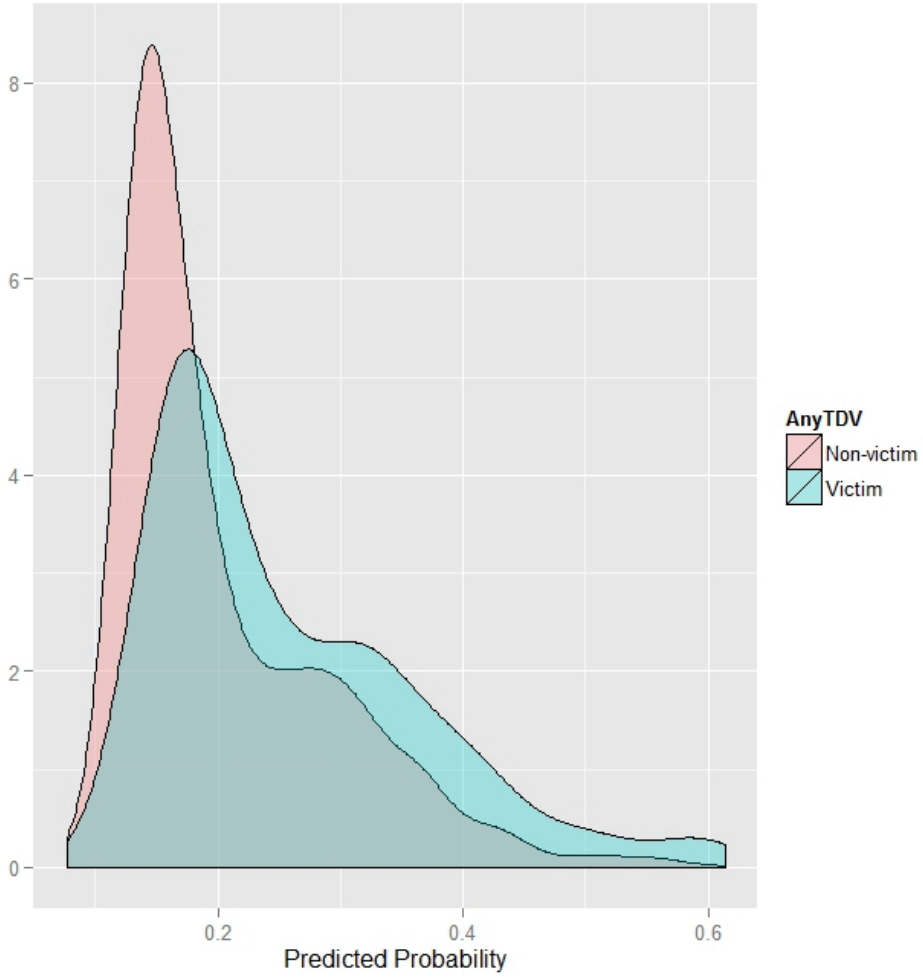

Figure 2.3. Predicted probability distribution for male participants of the probability of Wave 4 IPV victimization. Predicted probabilities were created using parameter estimates from the survey-weighted model for the full sample.

Table 2.5
Path Analysis: Female Participants

	Full Sample (*n*=2,331)		**Matched Sample (*n*=1,269)**	
	b (95% CI)	**p-value**	**b (95% CI)**	**p-value**
Dependent Variable: Wave 4 IPV Victimization				
W2 TDV victimization (b2)	-0.018 (-0.24, -0.17)	0.89	-0.017 (-0.17, 0.14)	0.86
Wave 3 IPV victimization (b1)	0.22 (0.12, 0.31)	<.001	0.24 (0.14, 0.34)	<.001
Propensity score (b3)	0.09 (-0.02, 0.23)	0.23	0.19 (0.07, 0.31)	.007
Dependent Variable: Wave 3 IPV Victimization				
Wave 2 TDV victimization (g1)	0.43 (0.31, 0.55)	<.001	0.35 (0.22, 0.48)	<.001
Propensity score (g2)	0.18 (0.06, 0.28)	.005	0.28 (0.18, 0.38)	<.001
Thresholds				
Wave 3 IPV victimization	0.37 (0.20, 0.61)	.003	0.23 (0.08, 0.39)	.014
Wave 4 IPV victimization	1.22 (0.92, 1.49)	<.001	0.75 (0.59, 0.93)	<.001
Indirect Paths				
Total (direct + indirect)	0.07 (-0.12, 0.25)	.53	0.07 (-0.08, 0.22)	.47
Total indirect	0.09 (0.05, 0.15)	.002	0.08 (0.04, 0.14)	.003

Full sample analyses also controlled for child maltreatment, race and family structure, and accounted for complex survey design effects (unequal probability of selection, clustering and post-stratification). Matched sample analyses were weighted to account for the 1:2 nearest neighbor match.

Table 2.6
Path Analysis: Male Participants

	Full Sample (*n*=1,605)		**Matched Sample (*n*=876)**	
	b (95% CI)	**p-value**	**b (95% CI)**	**p-value**
Dependent Variable: Wave 4 IPV Victimization				
W2 TDV victimization (b2)	-0.072 (-0.29, 0.15)	.59	-0.056 (-0.23, 0.12)	0.60
Wave 3 IPV victimization (b1)	0.30 (0.20, 0.42)	<.001	0.33 (0.22, 0.44)	<.001
Propensity score (b3)	0.11 (-0.01, 0.24)	.14	0.20 (0.07, 0.33)	.013
Dependent Variable: Wave 3 IPV Victimization				
Wave 2 TDV victimization (g1)	0.39 (0.17, 0.61)	.005	0.56 (0.37, 0.72)	<.001
Propensity score (g2)	0.15 (0.029, 0.31)	.074	0.04 (-0.09, 0.18)	.60
Thresholds				
Wave 3 IPV victimization	0.72 (0.45, 0.97)	<.001	0.82 (0.63, 1.03)	<.001
Wave 4 IPV victimization	0.71 (0.45, 0.97)	<.001	0.55 (0.37, 0.73)	<.001
Indirect Paths				
Total (direct + indirect)	0.045 (-0.16, 0.29)	.73	0.13 (-0.038, 0.30)	.21
Total indirect	0.12 (0.047, 0.21)	.016	0.18 (0.11, 0.28)	<.001

Full sample analyses also controlled for child maltreatment, race and family structure, and accounted for complex survey design effects (unequal probability of selection, clustering and post-stratification). Matched sample analyses were weighted to account for the 1:2 nearest neighbor match.

Discussion

In this sample, dating violence experienced during adolescence was directly associated with partner violence experienced approximately 5.5 years later, and was indirectly associated with partner violence experienced 12 years later through the experience of 5.5-year victimization. This study adds to the growing body of literature demonstrating the importance of TDV for understanding future risk for re-victimization, and extends prior work by examining how this risk is transmitted.

Using this same sample, Cui et al. (2013) also examined associations between Wave 2 dating violence and Wave 4 IPV victimization using propensity scores, and found that Wave 2 dating violence was significantly associated with Wave 4 IPV in both survey-weighted and propensity-matched samples. There are several potential reasons why their findings differ from those presented here. First, Cui et al.'s (2013) propensity score model only controlled for socio-demographic variables (e.g., gender, family structure), adolescent general aggression and child maltreatment. Thus, it is possible that the large number of additional confounders included in the present paper accounted for the relationship between Wave 2 TDV and Wave 4 IPV, and not dating violence itself. Secondly, their dating violence sample was much larger (n=1,158), suggesting that individuals who experienced victimization prior to the Wave 1 interview were not excluded (this exclusion is not specifically mentioned in their paper), which can lead to biased estimates (Rosenbaum, 1984). Finally, they used chronicity scores, and not prevalence scores, for both dating violence and Wave 4 IPV, which may also have led to the different findings.

While we had originally planned to examine mediation by depression and substance use, sample size limitations precluded us from investigating these pathways. However, even given our current findings, it is possible that depression and substance use are involved in continuing

risk for victimization, especially given the long period of time between measurement occasions. Lehrer et al. (2006) discuss that depression might be implicated in risk for victimization via the process of assortative mating (Krueger, Moffitt, Caspi, Bleske, & Silva, 1998). Specifically, adolescents experiencing depression are more likely to associate with deviant peers, and future mates are likely to be selected from within these deviant peer groups; thus, selected partners may have an increased likelihood to perpetrate violence (Lehrer et al., 2006, p. 274). While Lehrer et al. (2006) only considered depression in their study, the assortative mating theory may also extend to substance use, given associations between substance use and risky peer affiliations (La Greca, Prinstein, & Fetter, 2001; Prinstein, Boergers, & Spirito, 2001), and between substance use and partner victimization (Testa et al., 2003).

In addition to assortative mating, social learning theory—a theory used to understand risk for perpetration of IPV (e.g., Shorey, Cornelius, & Bell, 2008)—can also be used to understand risk for victimization through a consideration of the construct of self-efficacy (Bandura, 1977; Bandura, Adams, & Beyer, 1977). Efficacy expectations affect an individual's belief that he or she can successfully accomplish a behavior required to produce a desired outcome (Bandura et al., 1977), and Engstrom, El-Bassel, Go and Gilbert (2008) discuss how substance use might lower these expectations, leading to increased difficulty to leave or address abuse in a violent relationship. Assortative mating and social learning theory may also work in concert; for example, lowered efficacy expectations may make an individual more vulnerable to depression (e.g., Muris, 2002), which in turn may increase their affiliation with risky peers, and may culminate in increased risk for future partner victimization. While these theoretical mechanisms could not be empirically evaluated using the present sample, they are important topics for future research.

The use of propensity scores is a strength of the current study, allowing the inclusion of a

number of important pre-treatment covariates (including peer deviance and self-efficacy), and providing more robust support for the assumption of strongly ignorable treatment assignment (i.e., no unmeasured confounding); however, the complex survey design of the Add Health data did not allow for a straightforward application of propensity score matching. In order to address this, we ran all results on data that accounted for the survey design effects (i.e., clustering, post-stratification and unequal probability of selection), as well as on data that contained a matched treatment and control group. In the case of our bivariate analyses, these two methods resulted in slightly different frequencies, though this difference was typically small. A larger difference was seen in the significance patterns for bivariate associations, with survey-weighted results generally more conservative than matched sample results. These differences are not surprising, since sample weighting affects the parameter estimates (e.g., observed and expected cell frequencies), and both sample weighting and clustering increase the estimated variance (Chantala, 2006). In our multivariate analyses, there were also small differences in parameter estimates produced by the two methods, though the general pattern of results was similar, and the same final conclusion was reached. However, it is possible that in other analyses using this two-method solution, similar results would not be found (Korn & Graubard, 1995; Wang, Yu, & Lin, 1997), leading to difficult questions about the cause of this difference (e.g., the improved covariate balance resulting from matching? Or, biased estimates of population means in the matched, unweighted sample?), and so it is not an ideal solution. Rather, methods should be developed that account for the effects of matching within the survey framework, for example by building on work that discusses the inclusion of both design- and model-based weighting within model-based regression analysis (Little, 2007), or work that discusses sub-classifying on the propensity score as a way to incorporate complex survey design characteristics (Zanutto, 2006).

Finally, while we used propensity score matching to balance the covariate distribution

between the treatment and control groups, intended to meet the assumption of strongly ignorable treatment assignment given the covariates, matching was not performed on our mediator. Thus, while meditational analyses were conducted within a causal framework (Múthen, 2011), the lack of randomization on the mediator likely violates an assumption required for valid causal inference (namely, that there be no unmeasured mediator-outcome confounding, or, the assumption of sequential ignorability), and so a causal interpretation of the direct and indirect effect may not be appropriate, even given the temporality of assessments.[4] Instead, we consider the demonstrated effects risk factors for IPV victimization in adulthood. Future work should consider the issue of sequential ignorability when examining mediation, potentially by using propensity scores to match on the mediator (Jo, Stuart, MacKinnon, & Vinokur, 2011). It is also important to note that while propensity score matching aims to balance the observed covariates, it does not necessarily balance all unobserved covariates, as occurs in a true experimental design; thus, there may exist an unobserved confounder of the treatment-outcome relationship that accounts for the observed results. However, we selected a thorough list of observed covariates, and believe the assumption of ignorable treatment assignment in this sample is plausible; future work using sensitivity analyses will serve to explore this assumption (Rosenbaum, 2002).

Other limitations of our analyses pertain to the measurement of Wave 2 dating violence. The five dating violence questions measured relatively mild forms of psychological and physical aggression, and did not assess sexual victimization; it is possible that associations with Wave 4 IPV would be different if more severe items were included. Additionally, all five TDV questions were derived from the CTS2 (Straus et al., 1996), and so are focused on specific behaviors, and not the context within which the acts occurred, limiting a more nuanced investigation of the association of TDV with risk for future victimization. Finally, it is possible that some of the

[4] These topics (causality, mediation and structural equation models) are the focus of an active area of work; see Bollen and Pearl (2013) and Bullock, Green and Ha (2010) for an elaboration.

control group experienced TDV in the intervening years between the Wave 2 and Wave 3 assessments, which may attenuate our results.

Thus, while our findings provide a broad framework for understanding the longitudinal association of dating violence with future victimization, they should be supplemented with more detailed work in targeted samples, including work that uses more nuanced assessments of TDV (see Chapter Five for an example of such an assessment). However, the current findings underscore the public health importance of TDV, and the need for continued work on appropriate prevention and intervention. Understanding how experiences in early dating relationships translate into risk for re-victimization is an important task for future research.

References

Ackard, D. M., Eisenberg, M. E., & Neumark-Sztainer, D. (2007). Long-term impact of adolescent dating violence on the behavioral and psychological health of male and female youth. *Journal of Pediatrics, 151*, 476-481.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 194-215.

Bandura, A., Adams, N. E., & Beyer, J. (1977). Cognitive processes mediating behavioral change. *Journal of Personality and Social Psychology, 35*, 125-139.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301-330). New York, NY: Springer.

Breiding, M. J., Black, M. C., & Ryan, G. W. (2008). Chronic disease and health risk behaviors associated with intimate partner violence – 18 U.S. States/Territories, 2005. *Annals of Epidemiology, 18*, 538-544.

Brown, A., Cosgrave, E., Killackey, E., Purcell, E., Buckby, J., & Yung, A. R. (2009). The longitudinal association of adolescent dating violence with psychiatric disorders and functioning. *Journal of Interpersonal Violence, 24*, 1964-1979.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*, 550-558.

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner Abuse, 3*, 231-280.

Centers for Disease Control and Prevention [CDC]. (2012). Youth risk behavior surveillance – United States, 2011. *Morbidity and Mortality Weekly Report, 61*, 1-162.

Chantala, K. (2006). *Guidelines for analyzing Add Health data.* University of North Carolina at

Chapel Hill: Carolina Population Center.

Chantala, K., & Tabor, J. (1999, June). *Strategies to perform a design-based analysis using the Add Health data*. University of North Carolina at Chapel Hill: Carolina Population Center.

Cleveland, H. H., Herrera, V. M., & Stuewig, J. (2003). Abusive males and abused females in adolescent relationships: Risk factor similarity and dissimilarity and the role of relationship seriousness. *Journal of Family Violence, 18*, 325-339.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., … , Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23*, 260-268.

Crockett, L. J., Randall, B. A., Shen, Y., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the Center for Epidemiological Studies Depression Scale for Latino and Anglo adolescents: A national study. *Journal of Consulting and Clinical Psychology, 73*, 47-58.

Cui, M., Gordon, M., Ueno, K., & Fincham, F. D. (2013). The continuation of intimate partner violence from adolescence to young adulthood. *Journal of Marriage and Family, 75*, 300-313.

Dahlberg, L. L., & Krug, E. G. (2002). Violence – A global public health problem. In E. Krug, L. L. Dahlberg, J. A. Mercy, A. B. Zwi, & R. Lozano (Eds.), *World report on violence and health* (pp. 1-56). Geneva, Switzerland: World Health Organization.

Dube, S. R., Anda, R. F., Felitti, V. J., Edwards, V. J., & Williamson, D. F. (2002). Exposure to abuse, neglect, and household dysfunction among adults who witnessed intimate partner violence as children: Implications for health and social services. *Violence and Victims, 17*, 3-17.

Elliott, D. S., Ageton, S. S., & Huizinga, D. (1985). *Explaining delinquency and drug use.* Beverly Hills, CA: Sage Publications.

Engstrom, M., El-Bassel, N., Go, H., & Gilbert, L. (2008). Childhood sexual abuse and intimate partner violence among women in methadone treatment: A direct or mediated relationship? *Journal of Family Violence, 23*, 605-617.

Exner-Cortens, D. (submitted manuscript). Theory and teen dating violence victimization: Considering adolescent development. *Developmental Review.*

Exner-Cortens, D., Eckenrode, J., & Rothman, E. (2013). Longitudinal associations between teen dating violence victimization and adverse health outcomes. *Pediatrics*, *131*, 71-78.

Foshee, V. A., Benefield, T. S., Ennett, S. T., Bauman, K. E., & Suchindran, C. (2004). Longitudinal predictors of serious physical and sexual dating violence victimization during adolescence. *Preventive Medicine, 39*, 1007-1016.

Foster, H., Hagan, J., & Brooks-Gunn, J. (2004). Age, puberty, and exposure to intimate partner violence in adolescence. *Annals of the New York Academy of Science, 1036,* 151-166.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233-239.

Gómez, A. M. (2011). Testing the cycle of violence hypothesis: Child abuse and adolescent dating violence as predictors of intimate partner violence in young adulthood. *Youth & Society, 43*, 171-192.

Halpern, C. T., Oslak, S. G., Young, M. L., Martin, S. L., & Kupper, L. L. (2001). Partner violence among adolescents in opposite-sex romantic relationships: Findings from the National Longitudinal Study of Adolescent Health. *American Journal of Public Health, 91*, 1679-1685.

Halpern, C. T., Spriggs, A. L., Martin S. L., & Kupper, L. L. (2009). Patterns of intimate partner violence victimization from adolescence to young adulthood in a nationally representative sample. *Journal of Adolescent Health, 45*, 508–516.

Hitlin, S., & Elder, Jr., G. H. (2007). Agency: An empirical model of an abstract concept. *Advances in Life Course Research*, *11*, 33-67.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011, June). *MatchIt: Nonparametric preprocessing for parametric causal inference.* Retrieved from http://gking.harvard.edu/matchit

Howard, D., Qiu, Y., & Boekeloo, B. (2003). Personal and social contextual correlates of adolescent dating violence. *Journal of Adolescent Health, 33*, 9-17.

Hughes, J. N., Chen, Q., Thoemmes, F., & Kwok, O. (2010). An investigation of the relationship between retention in first grade and performance on high stakes tests in 3rd grade. *Educational Evaluation and Policy Analysis, 32*, 166-182.

Jiang, X., & Peterson, R. D. (2012). Beyond participation: The association between school extracurricular activities and involvement in violence across generations of immigration. *Journal of Youth & Adolescence, 41*, 362-378.

Jo, B., Stuart, E. A., MacKinnon, D. P., & Vinokur, A. D. (2011). The use of propensity scores in mediation analysis. *Multivariate Behavioral Research, 46*, 425-452.

Kaestle, C. E., & Halpern, C. T. (2005). Sexual intercourse precedes partner violence in adolescent romantic relationships. *Journal of Adolescent Health, 36*, 386-392.

Kim-Godwin, Y. S., Clements, C., McCuiston, A. M., & Fox, J. A. (2009). Dating violence among high school students in southeastern North Carolina. *The Journal of School Nursing, 25*, 141-151.

Kitzmann, K. M., Gaylord, N. K., Holt, A. R., & Kenny, E. D. (2003). Child witnesses to domestic violence: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71*, 339-352.

Knoble, N. B., Capaldi, D. M., Shortt, J. W., & Kim, H. K. (2012). *Risk factors for adolescent*

dating violence: A systematic review. Poster presented at the 14th Society for Research on Adolescence Biennial Meeting, Vancouver, BC.

Korn, E. L., & Graubard, B. I. (1995). Examples of differing weighted and unweighted estimates from a sample survey. *The American Statistician, 49*, 291-295.

Krueger, R. F., Moffitt, T. E., Caspi, A., Bleske, A., & Silva, P.A. (1998). Assortative mating for antisocial behavior: Developmental and methodological implications. *Behavior Genetics, 28*, 173.

La Greca, A. M., Prinstein, M. J., & Fetter, M. D. (2001). Adolescent peer crowd affiliation: Linkages with health-risk behaviors and close friendships. *Journal of Pediatric Psychology, 26*, 131-143.

Lee, S. S. (2011). Deviant peer affiliation and antisocial behavior: Interaction with monoamine oxidase A (MAOA) genotype. *Journal of Abnormal Child Psychology, 39*, 321-332.

Lehrer, J. A., Buka, S., Gortmaker, S., & Shrier, L. A. (2006). Depressive symptomatology as a predictor of exposure to intimate partner violence among US female adolescents and young adults. *Archives of Pediatrics Adolescent Medicine, 160*, 270-276.

Lewis, S. F., & Fremouw, W. (2001). Dating violence: A critical review of the literature. *Clinical Psychology Review, 21*, 105-127.

Lichter, E. L., & McCloskey, L. A. (2004). The effects of childhood exposure to marital violence on adolescent gender-role beliefs and dating violence. *Psychology of Women Quarterly, 28*, 344-357.

Little, R. J. (2007). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association, 99,* 546-556

Maas, C. D., Fleming, C. B., Herrenkohl, T. I., & Catalano, R. F. (2010). Childhood predictors of teen dating violence victimization. *Violence & Victims, 25*, 131-149.

Max, W., Rice, D. P., Finkelstein, E., Bardwell, R. A., & Leadbetter, S. (2004). The economic toll of intimate partner violence against women in the United States. *Violence and Victims, 19*, 259-272.

Muris, P. (2002). Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. *Personality and Individual Differences, 32*, 337-348.

Múthen, B. (2011). *Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus.* Retrieved from http://gseis.ucla.edu/academic-programs/education/aqm/aqm-resources/events-2010-2011/Muthencausalmediation.pdf

Múthen, L. K., & Múthen, B. O. (2012). Mplus: Statistical analysis with latent variables. User's guide. Retrieved from https://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r6_web.pdf

National Center for Injury Prevention and Control [NCIPC]. (2003, March). *Costs of intimate partner violence against women in the United States.* Atlanta, GA: Centers for Disease Control and Prevention.

Offenhauer, P., & Buchalter, A. (2011). *Teen dating violence: A literature review and annotated bibliography* (DOJ Document No. 235368). Washington, DC: Library of Congress.

O'Keefe, M., & Treister, L. (1998). Victims of dating violence among high school students: Are the predictors different for males and females? *Violence Against Women, 4*, 195-223.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth & Adolescence, 17*, 117-133.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative

strategies for communicating indirect effects. *Psychological Methods, 16*, 93-115.

Primack, B. A., Swanier, B. S., Georgiopoulos, A. M., Land, S. R., & Fine, M. J. (2009). Association between media use in adolescence and depression in young adulthood: A longitudinal study. *Archives of General Psychiatry, 66*, 181-188.

Prinstein, M. J., Boergers, J., & Spirito, A. (2001). Adolescents' and their friends' health-risk behavior: Factors that alter or add to peer influence. *Journal of Pediatric Psychology, 26*, 287-298.

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Raiford, J. L., Wingood, G. M., & DiClemente, R. J. (2007). Prevalence, incidence, and predictors of dating violence: A longitudinal study of African American female adolescents. *Journal of Women's Health, 16*, 822-832.

Rich, C. L., Gidycz, C. A., Warkentin, J. B., Loh, C., & Weiland, P. (2005). Child and adolescent abuse and subsequent victimization: A prospective study. *Child Abuse & Neglect, 29*, 1373-1394.

Rivara, F. P., Anderson, M. L., Fishman, P., Bonomi, A., Reid, R. J., Carrell, D., & Thompson, R. S. (2007). Healthcare utilization and costs for women with a history of intimate partner violence. *American Journal of Preventive Medicine, 32*, 89-96.

Roberts, T. A., & Klein, J. (2003). Intimate partner abuse and high-risk behavior in adolescents. *Archives of Pediatrics & Adolescent Medicine, 157*, 375–380.

Roberts, T. A., Klein, J. D., & Fisher, S. (2003). Longitudinal effect of intimate partner abuse on high-risk behavior among adolescents. *Archives of Pediatrics and Adolescent Medicine, 157,* 880-881.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton

University Press.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society Series A, 147*, 656-666.

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer-Verlag.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2,* 169-188.

Sherer, M. (2009). The nature and correlates of dating violence among Jewish and Arab youths in Israel. *Journal of Family Violence, 24,* 11-26.

Shorey, R. C., Cornelius, T. L., & Bell, K. M. (2008). A critical review of theoretical frameworks for dating violence: Comparing the dating and marital fields. *Aggression and Violent Behavior, 13*, 185-194.

Smith, P. H., White, J. W., & Holland, L. J. (2003). A longitudinal perspective on dating violence among adolescent and college-age women. *American Journal of Public Health, 93*, 1104-1109.

Spriggs, A. L., Halpern, C. T., & Martin, S. L. (2009a). Continuity of adolescent and early adult partner violence victimisation: association with witnessing violent crime in adolescence. *Journal of Epidemiology & Community Health*, *63*, 741–748.

Spriggs, A. L., Halpern, C. T., Herring, A. H., & Schoenbach, V. J. (2009b). Family and school socioeconomic disadvantage: Interactive influences on adolescent dating violence victimization. *Social Science & Medicine, 68*, 1956-1965.

Straus, M. A. (2004). *Scoring the CTS2 and CTSPC.* Retrieved from http://gauss.unh.edu/~mas2/CTS28a3.pdf

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict

Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283-316.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1-21.

Teitelman, A. M., Ratcliffe, S. J., Dichter, M. E., & Sullivan, C. M. (2008). Recent and past intimate partner abuse and HIV risk among young women. *Journal of Obstetric, Gynecologic & Neonatal Nursing, 37*, 219-227.

Testa, C. R., & Steinberg, L. (2010). Depressive symptoms and health-related risk-taking in adolescence. *Suicide and Life-Threatening Behavior, 40*, 298-305.

Testa, M., Livingston, J. A., & Leonard, K. E. (2003). Women's substance use and experiences of intimate partner violence: A longitudinal investigation among a community sample. *Addictive Behaviors, 28*, 1649-1664.

Tschann, J. M., Pasch, L. A., Flores, E., Marin, B. V., Baisch, E. M., & Wibbelsman, C. J. (2009). Nonviolent aspects of interparental conflict and dating violence among adolescents. *Journal of Family Issues, 30*, 295-319.

van Dulmen, M. H. M, Klipfel, K. M., Mata, M. D., Schinka, K. C., Claxton, S. E., Swahn, M. H., & Bossarte, R. M. (2012). Cross-lagged effects between intimate partner violence victimization and suicidality from adolescence into adulthood. *Journal of Adolescent Health, 51*, 510-516.

Vézina, J., Hébert, M., Poulin, F., Lavoie, F., Vitaro, F., & Tremblay, R. E. (2011). Risky lifestyle as a mediator of the relationship between deviant peer affiliation and dating violence victimization among adolescent girls. *Journal of Youth & Adolescence, 40*, 814-824.

Wang, S., Yu, M., & Lin, L. (1997). Consequences of analyzing complex survey data using inappropriate analysis and software computing packages. *Public Health, 111*, 259-262.

Wekerle, C., & Wolfe, D. A. (1998). The role of child maltreatment and attachment style in adolescent relationship violence. *Development and Psychopathology, 10*, 571-586.

Wight, R. G., Botticello, A. L., & Aneshensel, C. S. (2006). Socioeconomic context, social support, and adolescent mental health: A multilevel investigation. *Journal of Youth & Adolescence, 35*, 115-126.

Williams, T. S., Connolly, J., Pepler, D., Craig, W., & Laporte, L. (2008). Risk models of dating aggression across different adolescent relationships: A developmental psychopathology approach. *Journal of Consulting and Clinical Psychology, 76*, 622-632.

Zanutto, E. L. (2006). A comparison of propensity score and linear regression analysis of complex survey data. *Journal of Data Science*, *4*, 67-91.

CHAPTER THREE

MEASUREMENT OF TEEN DATING VIOLENCE: A COMPREHENSIVE REVIEW, PART 1

Measurement of Teen Dating Violence: A Comprehensive Review, Part 1

Deinera Exner-Cortens

Lydia Gill

John Eckenrode

Author Note

Deinera Exner-Cortens, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

Lydia Gill, Department of Human Development, Cornell University.

John Eckenrode, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

This research was supported in part by a Doctoral Foreign Study Award from the Canadian Institutes of Health Research.

Correspondence concerning this article should be addressed to Deinera Exner-Cortens,

CAMH Centre for Prevention Science, 100 Collip Circle, Suite 100, London, ON N6G-4X8.

Contact: dme56@cornell.edu

Abstract

Measurement of teen dating violence has substantial implications for our understanding of prevalence, correlates and outcomes; thus, the selection of a measure for a given research study is an important task. Currently, however, no comprehensive compendium presents teen dating violence measures with evidence of reliability and validity, or discusses strengths and limitations of these "evidence-based" measures. Such a document would aid researchers and practitioners in the selection of appropriate measures, and would also identify gaps in the literature, as well as directions for future research. Thus, this two-part comprehensive review presents teen dating violence measures that have been the focus of psychometric testing. Part 1 of this review presents behavior measures (i.e., measures that assess victimization and perpetration), while Part 2 presents attitude measures. We also review empirical literature that uses identified measures. In Part 2, we conclude by discussing the implications of this review for teen dating violence measurement, and also present several "promising" measures for future study.

Measurement of Teen Dating Violence: A Comprehensive Review, Part 1

Since the first studies documenting the existence of violence in teen dating relationships (Burcky, Reuterman, & Kopsky, 1988; Henton, Cate, Koval, Lloyd, & Christopher, 1983; Mercer, 1988; Roscoe & Callahan, 1985; Roscoe & Kelsey, 1986), a growing literature has emerged (Foshee & Reyes, 2011), focused on understanding the prevalence, correlates and outcomes of this aggression (e.g, Bandyopadhyay, Deokar, & Omar, 2010; Foshee & Matthew, 2007; Hickman, Jaycox, & Aronoff, 2004; Vagi et al., 2013). Nationally, approximately 10% of adolescents report being hit, slapped or physically hurt on purpose by a boyfriend or girlfriend in the past year (CDC, 2012), and approximately 30% report the receipt of psychologically aggressive behaviors in their lifetime (Halpern, Oslak, Young, Martin, & Kupper, 2001), with a smaller percentage reporting forced sexual activity (~1%; Wolitzky-Taylor et al., 2008). Studies have also shown that teen dating violence is often mutual, with individuals both using and receiving aggression within a romantic relationship (e.g., Giordano, Soto, Manning, & Longmore, 2010; Gray & Foshee, 1997; Miller et al., 2013; Orpinas, Hsieh, Song, Holland, & Nahapetyan, 2013). In some of the most prolific dating violence research, Foshee and colleagues have used data from two samples of students in North Carolina to report on the development of physical, sexual and psychological aggression perpetration (Foshee et al., 2009; Foshee, Bauman, Linder, Rice, & Wilcher, 2007; Reyes & Foshee, 2013) and physical victimization (Brooks-Russell, Foshee, & Ennett, 2013) in adolescent romantic relationships. Other work has demonstrated that risk and protective factors for dating violence span the social ecology (Exner-Cortens, Eckenrode, Bunge, & Rothman, Chapter Two; Foshee & Reyes, 2011; Knoble, Capaldi, Shortt, & Kim, 2012; Offenhauer & Buchalter, 2011; Sherer, 2009), and that dating violence victimization is longitudinally associated with multiple adverse health outcomes (Ackard, Eisenberg, & Neumark-Sztainer, 2007; Brown, Cosgrave, Killackey, Purcell, Buckby, & Yung,

2009; Exner-Cortens, Eckenrode, & Rothman, 2013; Roberts, Klein, & Fisher, 2003; Teitelman, Ratcliffe, Dichter, & Sullivan, 2008). While much of the early dating violence research occurred in the United States and Canada, newer research uses data from diverse regions, including Australia (Brown et al., 2009), Israel (Sherer & Sherer, 2008), Italy (Connolly et al., 2010b), Mexico (Antônio & Hokoda, 2009), New Zealand (Jackson, Cram, & Seymour, 2000), South Africa (Boafo, 2011), Spain (Fernández-Fuertes & Fuertes, 2010) and Thailand (Pradubmook-Sherer, 2009).

Measurement of Teen Dating Violence

This past work has contributed to an understanding of the epidemiology of teen dating violence and demonstrated that dating violence is a global public health problem; however, a question that has received less attention in the dating violence literature is how dating violence is measured, and what implications this might have for knowledge about prevalence, correlates and outcomes. In the early stages of teen dating violence research, where the goal was to provide an immediate, initial scope of the problem, researchers used measures that were not developed for or psychometrically tested in adolescent populations. Thirty years on, it seems important to assess the state of teen dating violence measurement, in order to assess gaps in the field, and provide directions for future research (Wekerle & Tanaka, 2010): good measurement is imperative to the accurate understanding of these complex interpersonal relationships.

The development and testing of new measures that better capture this construct was included in a recent funding solicitation on teen dating violence (NIJ, 2013). This funding proposal also specifically requested that new measures be compared with existing measures; however, no comprehensive review of teen dating violence measures exists, making this task difficult. Two Centers for Disease Control and Prevention (CDC) compendia of youth and intimate partner violence measurement list Foshee, Fothergill and Stuart's (1992) Acceptance of

Couple Violence scale, Foshee et al.'s (1996) physical and psychological perpetration and victimization Safe Dates scales, and Straus, Hamby, Boney-McCoy and Sugarman's (1996) Revised Conflict Tactics Scale (CTS2) (as cited in Dahlberg, Toal, Swahn, & Behrens, 2005 and Thompson, Basile, Hertz, & Sitterle, 2006). In their assessment of these four scales, the CDC compendia only provide information on internal consistency reliability, with no information on evidence of validity. Further, internal consistency reliability for the CTS2 is from Straus et al. (1996), a study using college undergraduates. Thus, it is not clear from these compendia whether there is evidence of reliability (for the CTS2) and validity (for all four scales) in adolescent populations.

Beyond these documents, a recent review of intimate partner violence research found that for adolescent samples, the most commonly used measure was the CTS, followed by the Safe Dates scales, Wolfe et al.'s (1994) Conflict in Relationships scale and Wolfe et al.'s (2001) Conflict in Adolescent Dating Relationships Inventory, though many studies also combined measures (as cited in Capaldi, Knoble, Shortt, & Kim, 2012). A content analysis of intimate partner violence assessments mentions the Dating Violence Questionnaire (Próspero, 2006, as cited in Hays & Emelianchik, 2009), and a recent review of the dating violence literature discusses several others, including the Attitudes about Aggression in Dating Situations scale, the Justification of Verbal and Coercive Tactics scale, the Relationship Violence Inventory, and the Dating Safety Questionnaire (Protivnak & McRoberts, 2011). However, information on the psychometric properties of these measures, necessary for researchers and practitioners choosing and comparing scales, was not provided in any of this past work; it is possible that many of these scales have no or poor evidence of reliability and validity, as is the case with many adult measures of intimate partner violence (Hays and Emelianchik, 2009). It is also possible that other tested measures of teen dating violence exist which were not cited in these prior papers, since

none specifically focused on reviewing all existing measures.

Assessing Measures

While a list of existing measures would be useful, researchers and practitioners also need to know how well a particular measure indicates the construct of interest. Ignoring this information is similar to providing a list of evidence-based programs without any presentation or assessment of the quality of evidence supporting program effectiveness. In this situation, researchers and practitioners would be unable to make an informed choice on the best program for their needs, just as without evidence of test psychometrics, researchers and practitioners are unable to make an informed decision about the best test. According to Kline (2000), a good measure exhibits high reliability, high validity, high discriminatory power and extensive norms (p. 24).

Reliability. Reliability refers to the consistency of a measure as examined by internal consistency or test-retest, for which recommended cut-offs exist to help assess whether or not a test is highly reliable. For test-retest reliability, Kline (2000) recommends a correlation coefficient of .70 if the test will be used for individual purposes, with a lower coefficient acceptable for group research purposes (though lower coefficients imply more measurement error). For internal consistency reliability, the test coefficient (e.g., Cronbach's alpha) should also ideally be > .70, but unlike in test-retest reliability, where coefficients should be as high as possible, the internal consistency coefficient should not be too high, as this indicates that test items may be too singular in nature, especially if the test is short (p. 30). Finally, although high reliability is important for accurate prediction (i.e., because "a test cannot correlate with anything more highly than it does to itself" (Kline, 2000, p. 29)), it is only a necessary, and not sufficient, requirement for validity.

Validity. Validity (i.e., whether the test is assessing the chosen construct) encompasses

several forms, including face validity, concurrent validity, incremental and differential validity, predictive validity, content validity and construct validity (Kline, 2000), though all types are not applicable to all tests. For example, although researchers often state that they used expert review to establish content validity, content validity requires that there is a definable content area (for example, if the test was intended to assess math knowledge in 10^{th} grade students, the defined content area might be the 10^{th} grade mathematics textbook). Having a specific and bounded content area may not be reasonable for many public health and social science topics, including teen dating violence, in which case expert review serves to establish face, and not content, validity. Face validity, however, does not establish the true validity of a test (Kline, 2000). Concurrent validity requires that a benchmark exist against which the new test can be assessed; the benchmark measure needs to be an established, high quality test, which may not exist in many fields, limiting the assessment of concurrent validity. Predictive validity requires that there exist a known and convincing criterion measure that the current test can predict, as do incremental and differential validity (incremental and differential validity also require the inclusion of multiple tests in the research study); thus, the establishment of these types of validity may be hindered by the lack of a strong criterion.

While each of these latter forms of validity are limited and not necessarily applicable in a given research study, construct validity, a concept first elaborated by Cronbach and Meehl (1955), does not have these same problems. As described by Cronbach and Meehl (1955), construct validity "must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured" (p. 282). Construct validity is a test of the theory of the construct (Maher & Gottesman, 2005; Nunnally, 1967; Smith, 2005a), and like any theory, requires detailed elaboration. Thus, the evaluation of a test's construct validity is defined by the researcher in terms of specific *a priori* hypotheses, with the goal of providing

evidence from multiple sources; factor analysis may form part of this evaluation if there are specific hypotheses about the expected structure (Cronbach & Meehl, 1955). Without these hypotheses, factor analysis is a tool of test construction (Kline, 2000); thus, confirmatory factor analysis (CFA) is more likely to serve as a test of construct validity than exploratory factor analysis (EFA), since the latter by definition does not lend itself to specific hypotheses. Hypotheses that explore convergent and discriminant validity are also necessary for the establishment of construct validity (Campbell and Fiske, 1959, p. 81).

Unlike reliability, then, which can be established in a given research study by examining statistical parameters, any given study only provides evidence of validity, and the validity of a test is based on the weight of this evidence (Smith, 2005b) – as noted by Cronbach and Meehl (1955), "…rejecting the null hypothesis does not finish the job of construct validation. The problem is not to conclude that the test 'is valid' for measuring the construct variable. The task is to state as definitely as possible the degree of validity the test is presumed to have" (p. 290). Given the "major task" of establishing construct validity (p. 296), as well as the subjective nature of the evidence, specific findings must be cited when evaluating and discussing the validity of a test, to allow for open dialogue and critique (Smith, 2005a).

Discriminatory power and norms. Discriminatory power refers to the variability in test scores, and the test's ability to discriminate among individuals (i.e., if every subject gets the same score on a particular test, it has no discriminatory power; Kline, 2000). Discriminatory power can be established through use of the delta statistic, for which Kline (2000) recommends a minimum value of .90 (a normal distribution has a delta value of .93). Together, reliability, validity and discriminatory power serve as indicators of the intrinsic quality of a given test. Norms, while also important to assessing overall test quality, are not a function of a particular test, but rather come from the administration of the test to relevant groups: scores from these

groups serve to establish norms, which are then used to aid in the interpretation of test scores in subsequent administrations (Kline, 2000, p. 40).

In sum, to assess the quality of a given measure, just as one might aim to assess the quality of a public health program, each of these characteristics (reliability, validity, discriminatory power and norms) can be examined, and evidence presented that supports (or refutes) claims of test quality.

Current Study

Since there exists no comprehensive review of teen dating violence measures, it is difficult to make definitive statements about the current state of measurement, and the needs of the field; knowledge on the full scope of teen dating violence measures with evidence of reliability and validity would help researchers and practitioners choose quality measures, as well as provide information on research gaps. Thus, the current paper offers a comprehensive review of teen dating violence measures, in order to provide a compendium of measures with evidence of reliability and validity. To supplement this assessment, this review also provides prevalence estimates obtained in empirical articles using these measures over the past 10 years.

Method

Search Strategy

To find relevant articles for this review, we followed several of the procedures recommended by Johnson and Eagly (2000), including 1) computer-based searches using online databases (e.g., PsycInfo, Web of Knowledge); 2) the ancestry approach (i.e., identifying studies from the reference lists of existing review articles); 3) the descendancy approach (i.e., identifying a small number of important historical works on the topic, and reviewing articles that cite these works); and 4) the invisible college technique (i.e., contacting key researchers in the field for any

unpublished work on the topic of interest) (pp. 502-503).[5] Steps taken to meet each of these procedures are described below.

Search 1: Measurement Articles

Search strategy. For the computer-based searches, we searched the following online databases for articles whose purpose was to provide evidence on the reliability and validity of a measure of teen dating violence: 1) Applied Social Sciences Index and Abstracts; 2) Directory of Open Access Journals; 3) ERIC; 4) Gale Virtual Reference Library; 5) Government Printing Office;[6] 6) Historical Abstracts; 7) PLOS ONE; 8) ProQuest Dissertations and Theses; 9) PsycInfo; 10) PubMed; 11) SAGE Reference Online; 12) Social Services Abstracts; 13) SocINDEX/Academic Search Premier; 14) Sociological Abstracts/GenderWatch/International Bibliography of the Social Sciences; 15) Web of Knowledge and 16) WorldCat. Search terms used were *dating violence* or *dating abuse* or *dating aggression* or *date fight** or *partner violence* or *gendered adolescent interpersonal aggression* AND *teen** or *adolescen** or *youth** AND *measure** or *scale** or *assess**. Following computer-based searches, we then reviewed the reference lists of 34 existing review articles (published from 1987-2012) for articles that were missed by the computer-based searches. For the invisible college technique, we contacted 33 researchers in the field to see if they had unpublished psychometric information for dating violence measures that had been used in their prior publications; no additional information was found in this search. The descendancy approach was not used for this search.

[5] Johnson and Eagly (2000) also recommend hand searching important journals in the field. We performed a modified version of this step, by creating an *a priori* list of 14 journals that were deemed relevant to the topic (e.g., *Psychological Assessment, Journal of Consulting and Clinical Psychology, Journal of Interpersonal Violence*). Following the computer-based and ancestry searches, we then checked if each journal was represented in the selected articles. If a journal was not included on this list, we re-searched the computer databases to make sure that the journal had been included in the catalog of at least one database. If a journal had not been included in this catalog, we would have then hand-searched that journal for relevant articles; however, using this procedure, we found that all 14 journals were represented by our selected databases, and so we did not perform any hand searches.

[6] Because the Government Printing Office searches often returned several hundred articles, and the majority of these articles did not appear relevant (e.g., Congressional hearings, meeting summaries), we initially randomly selected 12 search term combinations. In these 12 searches (1,885 returns), no articles were found for inclusion, and so the remaining 36 searches were not performed in this database.

Inclusion criteria. Searches were restricted to quantitative articles published in English, Spanish or French between January 1983 and June 2012. The lower bound for this search was chosen as the year the first article on teen dating violence was published (Henton et al., 1983). Searches were not restricted by geographic region or publication format. To be included, the primary purpose of the article had to be establishing the reliability and validity of a measure of teen dating violence (any aspect of teen dating violence – victimization, perpetration, attitudes, knowledge, and/or beliefs),[7] and the study needed to use a sample aged 18 or younger, or a sample that was in middle and/or high school. To keep the search focused, measures needed to assess violence between adolescent dating partners, and not non-partnered experiences of and/or attitudes toward aggression (e.g., sexual harassment, general sexual violence). Articles were excluded if they used a college or adult sample.

Review procedures. Three upper-level undergraduate research assistants, who did multi-hour training sessions and worked under the supervision of the first author [DE-C], performed initial article selection. Prior to completing searches, each research assistant performed two rounds of practice searches in one of the computer databases, and results were checked by and discussed with the first author during team meetings. To determine inclusion, the research assistants scanned article titles and abstracts to ascertain whether the article met inclusion criteria. If a decision could not be made, the research team member and the first author reviewed the full text. Questions about inclusion were also discussed at weekly team meetings. When the initial article pool was completed, the first author reviewed the full text of the final list of articles to confirm that exclusions were made where appropriate.

To assess reliability of the measurement search, each research assistant performed three to four computer-based searches originally done by another team member (for 11 total reliability

[7] In Appendix 3.A, there are two measures presented that only include reliability information. These were included because they were cultural adaptations of more robustly tested measures.

searches). In 10 of these 11 searches, there was an exact match in the articles selected. In the remaining search, one article was selected that the original team member had not included; however, this article was not a primary measurement article, but rather an article that presented information on a early version of a measure, and so would not have been included on our final list. Further, this article was picked up in a number of other searches; thus, the large number of databases and search terms used also supports the reliability of the search.

The initial computer-based search returned 9,765 articles, of which 296 were selected (including duplicates). After removing duplicates and reviewing the full-text for non-duplicate articles, 14 articles remained. Common reasons for exclusion during full-text review were age of sample, a focus on general aggression (as opposed to violence between dating partners), not focusing on measure validation (e.g., only presenting information on internal consistency reliability) and the use of qualitative methodology. From the ancestry search, an additional 14 articles were initially selected, of which 13 were subsequently excluded (for the same reasons as the computer-search articles), leaving one included article. In addition to these 15 articles (14 from the computer-based search and one from the ancestry approach), we included an additional three articles that were found during the empirical search (see below), and one article that was identified in our prior work on this topic. Thus, the total number of included measurement articles was 19. The 19 included articles presented evidence on six unique behavioral measures (Table 3.1) and seven unique attitude measures (see Part 2).

Table 3.1
Summary of Measures: Behaviors

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Type of violence	Items and scoring
CTS							
Cascardi et al. (1999)	Modified Conflict Tactics Scale [M-CTS]	2,320[b]	49.1	Multi-racial (56.3% Caucasian)[a]	Suffolk County, Long Island, NY	PV PP PsV PsP	Number of items: 18 Number of sub-scales: 2 Assessment period: Current or most recent relationship Response options: 5-point Likert-type scale (1=*never*, 5=*always*)
Nocentini et al. (2011)	Revised Conflict Tactics Scale [CTS2]	1,628[a,c]	50.9[a]	Italian (96% from Italian backgrounds) and Canadian (70% Euro-Canadian)	Lucca, Tuscany, Italy; Toronto and Kingston, Ontario, Canada	PP	Number of items: 9 Number of sub-scales: n/a Assessment period: Current or most recent relationship Response options: 5-point Likert-type scale (1=*never*, 5=*very often*)
CADRI							
Fernández-Fuertes et al. (2006)	Conflict in Adolescent Dating Relationships Inventory [CADRI] – Spanish	572	58.4[a]	Spanish	Salamanca, Spain	PV PsV SV PP PsP SP	Number of items: 35 Number of sub-scales: 5 Assessment period: Current or most recent relationship (past year) Response options: 4-point Likert-type scale (0=*never*, 3=*frequently, 6 times or more*)

(Table 3.1 continues)

(Table 3.1 continued)

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Type of violence	Items and scoring
CADRI							
Fernández-Gonzalez et al. (2012)	Conflict in Adolescent Dating Relationships Inventory [CADRI] – Short Form	St 1: 277 St 2: 365	St 1: 69.3 St 2: 59.2	St 1: Primarily Caucasian (85.9%) St 2: Multi-racial (34.7% Caucasian, 27.0% African)	Ontario, Canada	PV PsV SV PP PsP SP	Number of items: 10 Number of sub-scales: 5 Assessment period: Current or most recent relationship (past year) Response options: 4-point Likert-type scale (1=*never*, 4=*often, 6 or more times*)
Hokoda et al. (2006)	Conflict in Adolescent Dating Relationships Inventory [CADRI] – Spanish	307	62.5[a]	Mexican	Mexicali and Monterrey, Mexico	PV PsV SV PP PsP SP	Number of items: 34 Number of sub-scales: 5 Assessment period: Current or most recent relationship (past year) Response options: 4-point Likert-type scale (0=*never*, 3=*often*)
Jouriles et al. (2005) and Jouriles et al. (2009)	Relationship Violence Interview [RVI] – CADRI Physical Abuse, Threatening Behavior and Verbal/Emotional sub-scales	125	52.0	Multi-racial (39% Hispanic, 34% Caucasian, 18% African-American)	Southwestern USA (suburbs)	PV PsV	Number of items: 18 Number of sub-scales: 3 Assessment period: Past two weeks Response options: 5-point Likert-type scale (0=*never*, 4=*four or more*)

(Table 3.1 continues)

(Table 3.1 continued)

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Type of violence	Items and scoring
CADRI							
Wolfe et al. (2001)	Conflict in Adolescent Dating Relationships Inventory [CADRI]	St 1: 393 St 2: 1,019 St 3: 70 St 4: 26	St 1: 49.4[a] St 2: 55.0 St 3: n/a St 4: 50.0[a]	Primarily Caucasian (St 1: 81%; St 2: 79%)	Ontario, Canada	PV PsV SV PP PsP SP	Number of items: 35 Number of sub-scales: 5 Assessment period: Current or most recent relationship (past year) Response options: 4-point Likert-type scale (0=*never*, 3=*often, 6 times or more*)
Other – Behavior							
Lavoie & Vézina (2001)	VIolence faite aux Filles dans les Fréquentations à l'Adolescence [VIFFA]	708	53.2[a]	Primarily Caucasian (French Canadian)	Québec, Canada	PV PsV SV PP PsP SP[d]	Number of items: 29 Number of sub-scales: 4 Assessment period: Worst/most difficult dating relationship (past year) Response options: Psychological and sexual – 4-point Likert-type scale (1=*never*, 4=*more than 10 times*) Physical – dichotomous (yes/no)

(Table 3.1 continues)

(Table 3.1 continued)

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Type of violence	Items and scoring
Other – Behavior							
Leisen (1999)	Adolescent Partner Aggression Scale [APAS]	St 1: 25 St 2: 195 St 3: 212	St 1: 76.0[a] St 2: 59.0 St 3: 57.0	St 1: Multi-racial (24.0% Native Hawaiian, 20.0% Filipino, 16.0% Hispanic) St 2: Multi-racial (26.0% White, 18.8% Native Hawaiian, 12.0% Filipino) St 3: Multi-racial (40.5% Black, 36.7% Hispanic)	St 1: Oahu, HI and Boston, MA St 2: Oahu, HI St 3: Boston, MA	PV PsV SV PP PsP SP	Number of items: 32 Number of sub-scales: 3 (V), 2 (P) Assessment period: Ever and current Response options: Ever – 6-point Likert-type-scale (0=*never*, 6=*more than 20 times*) Current – Dichotomous ("check below if this has happened in your current dating relationship")
Murphy et al. (2012)	Tendency to Resist or End Abusive Dynamics [TREAD] and Warning Sign Behaviors [WSBs]	152[e]	100	Australian	Victoria, Australia	PsV	Number of items: TREAD – 19; WSB – 23 Number of sub-scales: TREAD – 3; WSB – 5 Assessment period: TREAD – n/a; WSB – Past few months Response options: TREAD – open-ended (with scoring rubric); WSB – 6-point Likert-type scale (0=*no*, 5=*very often*)

(Table 3.1 continues)

(Table 3.1 continued)

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Type of violence	Items and scoring
Other – Behavior							
Schultz & Jaycox (2008)	Recent Fearful Dating Experiences [RFDE][f]	900[a,g]	54.3[a]	Primarily Latino	Los Angeles, CA	PsV	Number of items: 11 Number of sub-scales: 1 Assessment period: Past 6 months Response options: 6-point Likert-type scale (1=*disagree strongly*, 6=*agree strongly*)

Abbreviations: P – physical; Ps – psychological; S – sexual; P – perpetration; V – victimization; St – Study

[a]Hand calculated

[b]Sample overlapped substantially with Slep et al. (2001) [see Part 2].

[c]Italian sample (*n*=704), Canadian sample (*n*=924)

[d]Victimization is reported only by girls, and perpetration only by boys.

[e]Murphy et al. (2012) also present some results from a convenience sample of 426 young adult women (aged 18-25), recruited online. Data from this sample were used to assess content validity.

[f]A title for this scale is not given by the authors, and so this title is adapted from the description

[g]Effective sample size was not explicitly stated by the authors. This sample size was calculated from Table 4 in Schultz and Jaycox (2008), and represents the number of students in the total sample (*n*=1,259) who had ever dated. Participants from this sample were also used in Edelen et al. (2009), Orlando et al. (2006) and Rayburn et al. (2007) [see Part 2].

Quality assessment. Two procedures were used to assess the quality of the 13 identified dating violence measures. First, a detailed quality assessment template was created; the template drew on assessment templates presented by Evers, Sijtsma, Lucassen and Meijer (2010); the Scientific Advisory Committee of Medical Outcomes Trust (2002); Shields, Gwaltney, Paty and Shiffman (2006); and Zaza et al. (2000). This template guided a review of the measure description, the sampling frame, reliability evidence, validity evidence, data analysis, data reporting, participant burden and cultural/language adaptations (see Appendix 3.F). Quality assessment templates were completed for each of the 19 included articles by the first author. Strength of correlations was assessed using Cohen's (1988) effect size descriptors, where .10 is considered a small association, .30 is considered moderate and .50 is considered large. In our summaries, types of validity assessed were taken from the author's description, and therefore these labels may not always conform to how validity was described in the Introduction. The descriptors of type of violence assessed were also taken from the authors' definition, and so may not be directly comparable between studies (e.g., some authors included threatening behaviors as part of physical violence, and others included it as part of psychological aggression). For Table 3.1, the number of scale items was sometimes difficult to determine, and so these numbers represent our best inference. Information on these articles is presented in Table 3.1; information summarizing the type(s) of reliability and non-face validity evidence presented is summarized in Appendix 3.A.

In addition to these indicators, we felt that knowing the typical prevalence rates obtained with these measures in different samples would contribute to the quality assessment (e.g., by providing a crude indicator of discriminatory power); thus, a second search was performed for empirical articles that used these 13 measures, in order to obtain a range of typical prevalence estimates.

Search 2: Empirical Articles

Search strategy. As in the measurement search, the recommendations of Johnson and Eagly (2000) were followed. For the computer-bases searches, we used PsycInfo. Because of the large number of dating violence articles (Foshee & Reyes, 2011), and the restricted goal of the search (finding empirical articles that used the selected measures), searches were modified in order to target the returns. Specifically, instead of using the search field boxes, the "Tests & Measures" box was used to specify the name of the measure. However, for searches where the measure was not specific to adolescent samples (e.g., Conflict Tactics Scale), the search field boxes were used, in order to refine the search. For these searches, we entered the following terms into the search field boxes: *teen** or *youth** or *adolescen** AND *dating violence* or *dating abuse* or *dating aggression* or *date fight** or *partner violence* or *gendered adolescent interpersonal aggression*. For every search, the full name of the scale was entered into the Tests & Measures box, and, if applicable, a search was also performed using the abbreviated measure title (e.g., Adolescent Partner Aggression Scale and APAS).

For the ancestry approach, reference lists of the 26 review articles published since 2002 were examined (see Inclusion Criteria); abstracts of relevant empirical articles identified from these reference lists were checked for possible inclusion if the empirical article hadn't been previously identified in the computer-based search. For the descendancy approach, we used Google Scholar and Web of Knowledge to review empirical articles citing the 13 measure validation articles. Both the English and the translated title were reviewed for non-English articles. The invisible college technique was not used in this search.

Inclusion criteria. Searches were restricted to empirical articles published between May 2002 and May 2012. The restricted date range was chosen because the purpose of this search was not to systematically review the empirical literature, but rather to obtain a representative sample

of quantitative articles, and a ten-year search period was deemed reasonable for this purpose. Articles were included if they: reported the results of quantitative research in a peer-reviewed publication; used one of the 13 identified measures; were written in English, French, or Spanish; for the behavior scales, reported the percent or raw number prevalence or incidence of sexual, physical, psychological or stalking victimization and/or perpetration by a dating partner, and not means; reported results stratified by sex or for one sex only; used a sample aged 18 or under, or middle and/or high school students; and had a response rate ≥30%. In addition, if the study was a program evaluation, prevalence/incidence needed to be reported at baseline if the treatment group was included in the prevalence estimate. Articles were excluded if they used a college or adult sample, or if they used only a few items from the Conflict Tactics Scale.

Review procedures. A multi-step process was used to review empirical articles. As in Search 1, three upper-level undergraduate research assistants performed all searches under the supervision of the first author, and all research team members received multi-hour training on the search process. Prior to performing searches, each research assistant did a practice exclusion test on articles assigned by the first author, and any errors were then reviewed and discussed. Following this initial practice exercise, each research assistant also performed test searches in two of the databases, and results were checked by and discussed with the first author during a team meeting. For the empirical search, titles and abstracts of potential articles were reviewed by one member of the research team, with the full text reviewed for all articles where an exclusion could not be made based on the abstract alone.

To assess reliability of the empirical search, each research assistant performed three to four searches originally completed by another team member (for 10 total reliability searches). In 9 of these 10 searches, there was an exact match in the articles selected. In one of the searches, there was one article pulled in the original search that was missed in the reliability test search.

Of the 946 returns from all searches, 89 non-duplicate articles were initially selected. Full-text review by the first author resulted in the exclusion of an additional 51 articles, which did not meet inclusion criteria (i.e., because of age of participants; publication date; only using a few items from the CTS; not using selected scales; measuring witnessing of violence instead of receipt or perpetration; not stratifying by sex; not being a quantitative study; reporting victimization and perpetration combined, or a total abuse score; reporting general adolescent peer victimization; and reporting only means and not raw frequency or percent prevalence). The final empirical sample contained 38 articles.

Data abstraction. Eight upper-level undergraduate research assistants, who received extensive training prior to beginning data abstraction, abstracted data from the 38 articles, under the supervision of the first author, and using a standardized template created for this project. The template guided a review of the measure used, study design and sample, primary outcome measure, study population demographics, and the prevalence/incidence of dating violence victimization, perpetration and/or attitudes. All articles were abstracted by two team members; team members completed abstractions separately, and then met to discuss findings and come to consensus about any discrepancies. Abstractions were also discussed at weekly team meetings, and all prevalence/incidence numbers were reviewed by the first [DE-C] and second [LG] authors. If the article provided only raw frequencies, prevalence numbers were calculated by the second author, and checked by the first author. If the article presented a range of estimates (e.g., for moderate and severe violence), we chose to report the range, rather than providing a midpoint estimate. If two articles reported on the same data from the same sample, we chose to report numbers from one article only,[8] but also included the reference for the other article in Appendices 3.B-3.E.

[8] Generally the article with the more comprehensible data. For example, if one article reported a prevalence range and the other reported a total prevalence number, we would select the latter.

Results

Measures

Conflict Tactics Scale (CTS). Two studies have examined the psychometric properties of the CTS in adolescent populations (Table 3.1). In the first study, Cascardi, Avery-Leaf, O'Leary and Slep (1999) used a non-representative sample of 2,320 high school students from Long Island, New York to explore the validity of the Modified CTS (M-CTS; Pan, Neidig, & O'Leary, 1994). The M-CTS is an 18-item, modified version of the CTS (Straus, 1979), created due to perceived limitations of the original CTS to distinguish between mild and severe physical violence, and concerns that some physical and psychological CTS items did not represent the true function of the action, leading to confusion about whether items tapped physical or psychological aggression (e.g., the item "threatened to hit/throw something at your spouse"). To address this, Pan et al. (1996) added four psychologically and two physically coercive acts to the 18-item CTS, based on interviews with couples, and also removed one original item. This 23-item scale had a four factor structure in their sample of military spouses.

For their adolescent population, Cascardi et al. (1999) included 17 of these 23 items, and added one additional item ("discussed things calmly"), for a total of 18 items; however, three of these items tapped conflict negotiation, and were not included in any analyses.[9] They also changed the reporting period to the entire length of the relationship (as opposed to the past year), in order to reflect the short duration of many adolescent dating relationships, and changed the response options from a frequency count to a Likert-type scale (Table 3.1). Prevalence on the 15 items assessing psychologically (e.g., "sulked/refused to talk") and physically (e.g., "kicked, bit, hit") aggressive acts ranged from 1.3-78.6% (perpetration) and 1.7-73.9% (victimization). Under

[9] Using data from a Spanish sample, two other studies have also presented information on the reliability of the 18-item adolescent M-CTS (Lozano, 2009; Muñoz-Rivas, Rodríguez, Gómez, O'Leary, & González, 2007). However, the sample used in these studies exceeded the eligible age range for the present review, and so they are not discussed here (participants were aged 16-26, and the sample included university students. Results were not stratified by age).

EFA (oblique rotation), a two-factor structure emerged for both males and females, with Factor 1 containing items assessing physical aggression (e.g., "kicked, bit, or hit with fist"), and Factor 2 containing items assessing psychological aggression (e.g., "insulted or sworn at partner"). However, the loading of items onto these factors differed for males and females, and for perpetration and victimization; for example, Factor 2 for male victims and perpetrators also contained mild physical items (e.g., "pushed, grabbed or shoved"), and for male perpetrators, Factor 1 only contained severe physical items (e.g., "threatened with knife or gun").[10] Reliability coefficients were not provided for the full scale or any of the sub-scales.

To assess convergent validity of the M-CTS, Cascardi et al. (1999) included variables that were empirically associated with dating violence (Attitudes Towards Interpersonal Violence (Riggs, 1990) and the Dominating and Jealous Tactics Scale (Kasian & Painter, 1992)), hypothesizing that the magnitude of association should range from .10 to .40, and that variables indicating jealous and controlling behaviors should correlate with M-CTS sub-scales more strongly than variables indicating attitudes toward dating violence. They also assessed associations between the adolescent M-CTS sub-scales and sub-scales created using traditional CTS scoring for physical and psychological aggression (Straus, 1979), in order to determine whether the new sub-scales represented the same latent construct. Cascardi et al. (1999) also used CFA, but only to confirm the EFA factor structure, and not to support any specific hypotheses about the construct. Finally, Cascardi et al. (1999) state that there exists "preliminary support for the face validity of CTS items for adolescents" (p. 547), with support presumably coming from their pilot work, which included expert review. Concurrent validity was not assessed due to the lack of a benchmark measure (Cascardi et al., 1999).

The fit of the CFA model was better for females than males, for both perpetration and

[10] In this sample, results for males were presented for current daters only, because of differences in the factor structure for current and recent daters (results for females are for current and recent daters combined), suggesting that the same M-CTS scoring procedure may not be appropriate for male current vs. recent daters.

victimization. Using factor scores generated from the adolescent M-CTS CFA factors, as well as from traditional CTS scoring criteria (Straus, 1979), Cascardi et al. (1999) found strong correlations between these two scoring methods for male and female perpetration (all $r \geq .68$) and female victimization (all $r \geq .63$). For male victimization, correlations with the traditionally scored CTS physical aggression sub-scale were small to moderate in nature (r=.22 and r=.36 for mild physical/psychological and severe physical M-CTS CFA factors, respectively), but were stronger between CTS psychological aggression and the mild physical/psychological M-CTS CFA factor (r = .76). Correlations with attitudes and jealous/controlling behaviors were small to moderate in nature for perpetration (between .10 and .34), and associations between the physical aggression sub-scales and jealous/controlling behaviors were generally larger in magnitude than correlations with attitudes, as hypothesized (correlations with the psychological aggression sub-scales and attitudes were not reported). For victimization, correlations between physical and psychological aggression and the receipt of jealous/controlling behaviors were moderate (r range .25-.38 for M-CTS CFA factors), and were much smaller in magnitude for attitudes (r range .00-.25 for M-CTS CFA factors), again supporting hypotheses. The association between severe physical victimization and jealous/controlling behaviors was not reported for males. In general, then, these associations met the stated hypotheses, with the majority of correlations in the range .10-.40, though some associations between victimization and attitudes were smaller than .10 (no associations were greater than .40). Correlations with jealous and controlling behaviors were also typically larger in magnitude than correlations with attitudes, and most CFA sub-scales correlated strongly with traditional CTS physical and psychological aggression sub-scales.

Nocentini et al. (2011) examined the measurement invariance of the physical perpetration sub-scale of the CTS2 in a non-representative sample of Canadian and Italian adolescents (Table 3.1). Although the samples were similar in age and dating history, there was a large difference in

parental education between the two samples (Italian: 17% of parents reported post-secondary education; Canadian: 57% reported some post-secondary education). Pilot work on this scale was done using a different sample of Canadian participants, and as a result, two items were modified, several others combined, and two items added ("spitting"; "pulling hair or scratching"), for a final 9-item scale; six of the items represent minor physical acts, and three represent severe physical acts. Modifications made by Cascardi et al. (1999) to reporting period and response scale were also adopted here. For the Italian sub-sample, the measure was translated and back-translated.

In the Italian sample, item endorsement ranged from 5.7-14.1%, and in the Canadian sample, 4.7-18.9%. There were significant differences between the Italian and Canadian samples on two items ("pushing, grabbing, or shoving" and "throwing, smashing, hitting, or kicking an object"; the former was endorsed by more Canadian participants, and the latter by more Italian participants). To assess construct validity, Nocentini et al. (2011) used confirmatory factor analysis, and hypothesized that a one-factor solution would fit the data better than a two-factor solution, given that their community-based participants were unlikely to differentially use mild vs. severe physical actions against a dating partner.

Nocentini et al. (2011) also hypothesized that there would be substantial invariance by both country and gender. For all groups (Canadian males, Canadian females, Italian males and Italian females), a one-factor solution was found to fit the data; the fit was very good in all samples, although the solution was simpler (i.e., no correlated errors) and stronger (i.e., larger factor loadings) for male participants. On this one-item scale, internal consistency reliability was good for Italian males (α=.94) and Canadian males (α=.91) and females (α=.82), but was much lower for Italian females (α=.64). While three of these four values exceeded Kline's (2000) recommended cut-off of .70, the values for male participants were also quite high, suggesting

that the scale may be too narrowly focused. When looking at measurement invariance, Nocentini et al. (2011) found partial invariance for gender within country, and full invariance between countries, concluding that the partial invariance of the model by gender within each country demonstrates the importance of assessing male and female aggression separately.

Findings from these two studies provide some overlap regarding the reliability and validity of the CTS in an adolescent population, and both generally supported their construct validity hypotheses. Although these studies used different versions of the CTS, the items on Cascardi et al.'s (1999) physical perpetration sub-scale are similar in content to the items on Nocentini et al.'s (2011) physical perpetration scale. However, there were some differences. For example, there are four items on Nocentini's CTS2 adaptation that are not included on Cascardi's M-CTS adaptation, and several items from the M-CTS adaptation are combined on the CTS2 adaptation. It is also not entirely clear how these scales should be scored, as the Cascardi et al. (1999) CFA solution allowed for item cross-loadings, and both studies required correlated errors to find adequate fit of the CFA model to the data (which also suggests that similarities between some items are not captured properly by current items). Also, only Cascardi et al. (1999) provides information on victimization items, and Cascardi et al. (1999) did not provide any reliability information. Finally, neither study provides specific information on discriminatory power, although prevalence reported by Cascardi et al. (1999) suggests potentially poor discriminatory ability for some M-CTS items. Thus, while the CTS, in some revised form, is promising for use with adolescents, evidence is preliminary, and it is not clear what version of the test should be used, or how it should be scored.

Conflict in Adolescent Dating Relationships Inventory (CADRI). The CADRI was developed by Wolfe et al. (2001) to assess victimization and perpetration of psychological, physical and sexual dating violence; however, in the initial validation paper, only the

perpetration items were examined for reliability and validity.[11] In their paper, Wolfe et al. (2001) used four samples of Canadian adolescents – the first was used to refine item selection; the second to determine scale structure; the third to establish test-retest reliability and partner agreement; and the fourth to establish construct validity (Table 3.1). Items on the CADRI were developed using several adult measures (e.g., the CTS, the Psychological Maltreatment of Women Inventory), literature reviews, expert panels, and a pilot study, with item content further refined in Study 1. In order to help clarify the context under which actions occurred, the CADRI instructions prompt the participant to report on any violence that occurred within dating relationships in the past year while they were having an argument with their dating partner. The CADRI also includes a detailed screen to establish which dating partner the participant will reference when answering behavioral items; screen items were developed with the use of focus groups.

Using representative, Study Two data, Wolfe et al. (2001) found that prevalence on the 25 individual CADRI items ranged from 2.2% ("I forced him/her to have sex when he/she didn't want to") to 53.7% ("I spoke to him/her in a hostile or mean tone of voice"), with much higher prevalence for the verbal/emotional abuse items (range: 25.6-53.7%, compared to 2.2-20.6% for all other items). In this sample, Wolfe et al. (2001) hypothesized a second-order factor structure, where the first-order factors represented five types of abusive behavior (verbal/emotional, physical, sexual, relational aggression, and threatening behavior) and the second-order factor represented latent abuse. Differences in the model by age and sex were also examined, but no specific hypotheses were presented regarding potential differences.

As hypothesized, the first-order model was comprised of five factors: threatening behavior (e.g., "I threatened to hurt him/her"; 4 items, α=.66), relational aggression (e.g., "I

[11] The CADRI also contains items tapping positive conflict resolution behaviors, but these were not tested by Wolfe et al. (2001).

spread rumours about him/her"; 3 items, α=.52), physical abuse (e.g., "I pushed, shoved or shook him/her"; 4 items, α=.83), sexual abuse (e.g., "I forced him/her to have sex when he/she didn't want to"; 4 items, α=.51), and verbal/emotional abuse (e.g., "I said things just to make him/her angry"; 10 items, α=.82); internal consistency values for sexual abuse and relational aggression were quite low. The second-order latent abuse model was comprised of the five first-order factors (α=.83). The fit of this second-order structure was adequate, but was improved with the addition of several correlated errors.

Examining this second-order structure by age and sex, Wolfe et al. (2001) found that it did not fit well for all groups; for example, the loading for sexual abuse onto the latent abuse factor was much higher for Grade 9 students (λ=0.57) than Grade 11 students (λ=0.21). Thus, Wolfe et al. (2001) also tested a restricted second-order structure, with only three behavior types at the first level (physical abuse, threatening behavior, and verbal/emotional abuse). This restricted model had adequate fit for all groups (males and females; Grades 9, 10 and 11). However, even in this restricted model, several of the internal consistency reliabilities for the first-order factors were < .70, with only verbal/emotional abuse, physical abuse and the second-order abuse factor consistently exceeding the cut-off; internal consistency reliabilities were generally higher for males than females. Over all the sub-samples, internal consistency reliabilities for relational aggression ranged from .16-.69, for sexual abuse, .36-.59, and for threatening behavior, .54-.73.

Study Three examined test-retest reliability, as well as dating partner agreement about acts perpetrated, in a smaller sample of adolescents (Table 3.1); relational aggression items were not examined in this study. Test-retest reliability exceeded .70 for the verbal/emotional abuse sub-scale and the restricted second-order abuse factor, but was weaker for other sub-scales (physical abuse=.64; threatening behavior=.58; full abuse second-order factor=.68), with the

lowest agreement for the sexual abuse sub-scale (.28). Correlation coefficients for partner agreement were moderate to strong (*r* range .29-.64) for all sub-scales except threatening behaviors when the perpetrator was female (*r*=0.20). Physical abuse could not be assessed during the partner agreement task due to the low number of endorsements. Overall, then, reliability evidence was strongest for physical abuse, verbal/emotional abuse and the restricted second-order abuse factor; however, internal consistency reliability for sexual abuse, relational aggression and threatening behaviors, as well as test-retest information for sexual abuse and threatening behaviors, suggested limited reliability for these sub-scales. There were also potentially important differences in reliability by age and sex of sample, a finding which deserves further study. Given these results, Wolfe et al. (2001) recommend that the restricted form of the CADRI (excluding sexual and relational abuse) is the most reliable form of the instrument (p. 287), and may be most appropriate for exploring age and sex differences across time.

Construct validity was explored using a community sample of adolescents (Table 3.1). Wolfe et al. (2001) had 26 dating couples participate in a video-taped interaction task, which was coded by observers for mentions of CADRI items (e.g., spreading rumours). The scores from coded interactions were then correlated with the youth's overall (total and restricted) self-report perpetration score. For males, these correlations were moderate and significant (*r*=.44 and .43, respectively), while for females, they were not; Wolfe et al. (2001) suggest this was due to the small sample size (*n*=26) and limited behavior sampling (p. 286).

Fernández-Fuertes, Fuertes and Pulido (2006) translated the CADRI into Spanish, and investigated its psychometric properties in a sample of Spanish youth (Table 3.1). In this sample, Fernández-Fuertes et al. (2006) examined both the perpetration and victimization items, and used EFA to assess a five-factor structure (a second-order structure was not explored); however, since

they were attempting to verify if the Wolfe et al. (2001) structure held in their data, CFA would have been more appropriate for this task. No information was given on the translation process, and no specific validity hypotheses were explored.

Using EFA (orthogonal rotation), they initially found a six-factor structure, but forced the solution to five factors, in order to compare the structure to Wolfe et al. (2001). In this solution, some of the items loaded onto different factors than in the Wolfe et al. (2001) solution, with the threatening behavior sub-scale showing the least similarity. Only the item "I threatened to hit him/her or throw something at him/her" loaded onto the threatening behavior scale in both the Wolfe et al. (2001) and Fernández-Fuertes et al. (2006) studies; the remaining three items from the Wolfe et al. (2001) threatening behavior scale loaded on relational abuse, verbal/emotional abuse, and cross-loaded on relational and physical abuse. Internal consistency coefficients for the five Fernández-Fuertes sub-scales were similar to those reported in Wolfe et al. (2001), but were lower for the threatening behavior (α=.53 vs. α=.66, respectively), physical abuse (α=.73 vs. α=.83, respectively) and verbal/emotional abuse (α=.78 vs. α=.82, respectively) sub-scales. As in Wolfe et al. (2001), however, only the physical and verbal/emotional abuse coefficients exceeded .70. Fernández-Fuertes et al. (2006) also presented an EFA solution for the victimization items, which was forced to a five-factor solution. Here, too, there was an issue with the threatening behavior sub-scale (only two items loaded on this scale; the other items they expected to load on this factor loaded on sexual abuse and verbal/emotional abuse). Internal consistency reliabilities for the victimization sub-scales were low (threatening behavior α=.51, sexual abuse α=.56) to acceptable (relational aggression α=.73, physical abuse α=.76, verbal/emotional abuse α=.79). Means on the five perpetration and victimization scales indicated variability in responses, although they also indicated that the response distribution was likely positively skewed, limiting discriminatory power.

Hokoda et al. (2006) also translated the CADRI into Spanish, and tested its reliability in a non-representative sample of Mexican high school adolescents (noting that not all Mexican adolescents go to high school, and so this is a somewhat selective sample). They did not assess the factor structure, but calculated reliability based on the structure presented by Wolfe et al. (2001). To translate the CADRI, Hokoda et al. (2006) performed translation and blinded back-translation, discussed the measure with professionals who work with Mexican youth, and piloted the measure with Mexican adolescents, in order to get feedback and improve unclear questions. During this process, they lowered the reading level to 4th grade (from the original 12th grade) to improve translation success, and also deleted one item ("I forced my dating partner to have sex"), due to concern about reaction to this item, as well as mandated reporting issues. Hokoda et al. (2006) report that some of the physical aggression items were difficult to translate, due to the use of colloquial terms (e.g., "break up").

For male perpetration, two-week test-retest reliability exceeded .70 for the relational aggression, physical abuse and verbal/emotional abuse sub-scales, as well as the restricted abuse total scale. Test-retest reliability was lower for male victimization, with only the reliability coefficient for relational aggression exceeding .70. For female perpetration and victimization, test-retest reliability exceeded .70 for the physical abuse and verbal/emotional abuse sub-scales, and the restricted abuse total scale; for perpetration, reliability also exceeded .70 for the threatening behavior sub-scale. Internal consistency reliability was low to adequate for male and female perpetration and victimization, with the lowest reliability for male threatening behavior perpetration (α=.39), male sexual abuse victimization (α=.43), and female sexual abuse perpetration (α=.37) and victimization (α=.39). Issues with reliability of the sexual abuse and threatening behavior sub-scales, as well as the relational aggression sub-scale for females, reflect findings by Wolfe et al. (2001) of the more robust reliability of the physical and verbal/emotional

abuse scales. No validity hypotheses were examined in this paper and no information on discriminatory power was presented. The authors also did not assess whether a five-factor structure was appropriate in this population, although they discuss that understanding of concepts may differ across culture, such that exploration of this structure seems warranted.

Fernández-González, Wekerle and Goldstein (2012) evaluated a 10-item, short-form of the CADRI perpetration items (CADRI-S) in two samples (Table 3.1). Item selection for the CADRI-S was guided by six criteria involving statistical and content factors (pp. 38-39); using these criteria, two items were selected from each original sub-scale, so that each latent construct was equally represented. If several items were equally viable for a particular sub-scale, Fernández-González et al. (2012) chose the item that was more frequently endorsed, and so this scale may be capturing more normative interactions than the full CADRI. Time to complete the CADRI-S is much shorter than time required to complete the full CADRI (3-5 minutes vs. 10-15 minutes, respectively).

For the entire 10-item scale, internal consistency reliability, as indicated by Cronbach's alpha, was .85, and in Study 1 (high school students), was slightly higher for males than females (α=.90 vs. α=.69, respectively), which reflects findings on the full CADRI. In Study 2 (child protective services (CPS)-involved youth), Cronbach's alpha was higher for females (α=.91) than males (α=.68), though the overall reliability was still good (α=0.81).

To assess construct validity, Fernández-González et al. (2012) used CFA to test a hypothesized second-order factor structure, finding a well-fitting second order-structure in both samples and replicating the structure presented by Wolfe et al. (2001). However, the magnitude of some item loadings differed between the CPS sample (Study 2) and the high school sample (Study 1), and loading of relational abuse onto the second-order abuse factor was much lower in the CPS sample (λ=0.54, CPS vs. λ=0.97, high school). Due to sample size limitations, CFA was

not stratified by sex in either Study 1 or Study 2 (all other results were presented for the total sample, and stratified by sex); however, given the results of Wolfe et al. (2001), it will be important to investigate the gender and age invariance of the CADRI-S in a subsequent study. Concurrent validity was assessed by correlating scores on the CADRI-S with scores from the full CADRI. In both Study 1 and 2, correlations were strong for males and females (*r* range: .71-.96), except for the correlation between the physical abuse short and full sub-scales for male CPS participants, which was moderate (*r*=.40).

In Study 2, Fernández-González et al. (2012) also examined correlations between the CADRI-S and full CADRI and bullying, to provide evidence of convergent validity; they did not hypothesize about the strength of the correlation between these measures, but rather wanted to determine if the pattern of correlations was the same for the CADRI-S and full CADRI. For these correlations, they found mixed results, with a slightly different pattern of associations for the short and full forms in males and females. However, these associations were generally similar in magnitude, except for the association between bullying perpetration and the physical abuse sub-scale for males, which was negative and non-significant for the CADRI-S (*r*=-.05), and positive and significant for the full CADRI (*r*=.21, $p < .05$). Using Study 2 data, they also examined predictive validity, by examining associations between the short and full CADRI at baseline and the full CADRI at six-month follow-up[12]; in the total sample, correlations were moderate to strong for both the CADRI-S and full CADRI, with the lowest correlation for the sexual abuse sub-scale (*r,* total sample, CADRI-S: .22; *r*, total sample, full CADRI: .27). When stratifying by sex, the sexual abuse sub-scales for both males and females, and the relational abuse sub-scale for males only, also had weaker associations with the full CADRI at 6-month follow-up (*r*, range=.22-.26).

[12] Fernández-González et al. (2012) refer to this as predictive validity, but were looking at the correlation between CADRI scores at baseline and six-month follow-up, and thus it may be more accurate to consider it a form of test-retest reliability.

Finally, to assess discriminatory power compared to the full CADRI, Fernández-González et al. (2012) calculated sub-scale and total scale prevalence rates for both the CADRI-S and full CADRI. The total scale prevalence rate was significantly lower on the CADRI-S than the full CADRI ($p < .001$), and was also lower on all sub-scales; on the CADRI-S, sub-scale prevalence rates ranged from 10.9-64.5% (Study 1) and 6.7-63.0% (Study 2), compared to 17.8-92.4% (Study 1) and 12.5-89.5% (Study 2) on the full CADRI. As in the original validation, prevalence of verbal/emotional abuse was much higher than prevalence on the other sub-scales.

The CADRI-S is an improvement over the non-tested, modified short scales often used in the field due to time limitations; unlike these scales, the CADRI-S covers the same content areas as the original CADRI and has preliminary evidence of reliability and validity. Further, this study demonstrated the use of the CADRI-S with both high-risk (CPS) and general population youth. However, more work is needed to establish its utility in clinical settings. Also, because of the lower sensitivity of this measure (compared to the full CADRI), Fernández-González et al. (2012) suggest that it would be more useful in studies aiming to understand associations between dating violence and other variables, and not studies where the goal is to establish prevalence.

Jouriles, Mcdonald, Garrido, Rosenfield and Brown (2005) also investigated a short form of the CADRI, with an altered recall period.[13] Using the physical violence (4 items) and threatening behavior (4 items) CADRI victimization sub-scales, Jouriles et al. (2005) created the Relationship Violence Interview, which uses four cumulative, two-week assessment periods (spaced two weeks apart) and is interviewer-administered; the cumulative assessments of the RVI are used to improve recall, with Jouriles et al. (2005) hypothesizing that the cumulative assessment strategy would result in higher prevalence than a one-time recall period. The RVI also contains follow-up prompts to assess the participant's perception of endorsed behaviors

[13] Although Jouriles et al. (2005) state that their study was not intended to present a new measure, it is discussed here because of its implications for CADRI recall period. Psychometric information on this measure is also provided in Jouriles, Garrido, Rosenfield and McDonald (2009).

(e.g., if they thought the partner was playing around). For comparison purposes, Jouriles et al. (2005) also administered a one-time, two-month recall version of the CADRI, though this reporting period is different from the typical 12-month CADRI recall period, and both the RVI and two-month CADRI were interviewer-administered, as opposed to a self-administered, questionnaires. While the individual RVI assessments take the same amount of time as a typical CADRI administration (10-15 minutes), the repeated nature of these assessments makes them somewhat more burdensome.

In Jouriles et al. (2005), internal consistency reliability, as assessed by Cronbach's alpha, of the two RVI sub-scales was .71 (physical aggression) and .70 (threatening behavior). The reliability for the physical aggression sub-scale was similar to the reliability found by Fernández-Fuertes et al. (2006) and Hokoda et al. (2006), but threatening behavior reliability was higher (.70, compared to .51 and .57, respectively). As hypothesized, the cumulative prevalence over the two-month period was significantly higher than the one-time prevalence rate (48.0% vs. 27.2% for physical aggression and 48.8% vs. 24.8% for threatening behaviors, respectively). However, individual two-week assessment period prevalence rates were similar to the one-time, two-month assessment period prevalence (e.g., for two-week assessment period one, the prevalence was 25.6% for physical abuse and 28.0% for threatening behavior).

To assess convergent validity, Jouriles et al. (2005) examined correlations between the RVI and trauma symptoms (intrusive thoughts, avoidance, hyper-arousal) and relationship anxiety, hypothesizing that the RVI sub-scale scores would be more highly associated with these variables than the CADRI (one-time) sub-scale scores. This hypothesis was partially supported. Both the RVI-C (cumulative) physical and threatening sub-scales were significantly correlated with trauma and relationship anxiety, while the CADRI subscales were not. However, the difference between CADRI and RVI-C threatening behavior correlation coefficients was not

significant, such that RVI-C threatening behavior was not more strongly associated with trauma and relationship anxiety than CADRI threatening behavior. For the physical aggression sub-scale, though, RVI-C scores were more strongly associated with trauma and relationship anxiety than CADRI scores (similar findings emerged when both sub-scale scores were entered into a regression model). Jouriles et al. (2005) also examined correlations between the CADRI sub-scales and the RVI-C sub-scales, and found that the RVI-C physical abuse and threatening behavior sub-scales were strongly correlated ($r \sim .50$) with the respective CADRI sub-scales.

Jouriles et al. (2005) also hypothesized that the one-time CADRI assessment would detect more recent aggression experiences, as opposed to more distant ones, and found that this was the case; for both the physical abuse and threatening behavior sub-scales, more recent aggression was more likely to be reported on the CADRI (determined by examining when aggression was last reported on the RVI). Specifically, aggression that was reported on the RVI more than two weeks prior had approximately a 50% lower chance of being reported on the CADRI. Jouriles et al. (2005) also conducted several follow-up analyses on the RVI and CADRI by participant sex, interviewer-participant sex concordance, intact vs. ended relationship status, and reference period. In regards to participant sex specifically (as Wolfe et al.'s (2001) research indicated that the CADRI may not perform equally well across the sexes), Jouriles et al. (2005) found that there were no differences in the pattern of reported results for males and females, except for the association between physical aggression and trauma symptoms, which was qualified by a significant interaction, such that there was a positive association between CADRI scores and trauma symptoms for female participants, and a negative association between CADRI scores and trauma symptoms for male participants. Other secondary analyses demonstrated little to no difference on the pattern of results when considering the other moderator variables.

Using this same sample, Jouriles et al. (2009) examined the physical (4 items) and

verbal/emotional abuse (10 items) sub-scales of the CADRI, also using a two-week recall period and cumulative assessments; only results for verbal/emotional abuse are discussed here, since results for the RVI physical aggression sub-scale were presented in Jouriles et al. (2005). The internal consistency reliability for the RVI-C verbal/emotional sub-scale was .81 (Cronbach's alpha). As in Jouriles et al. (2005), the cumulative prevalence of verbal/emotional abuse was much higher than the one-time prevalence (96% vs. 91%, respectively), though for both assessment periods, this high prevalence suggests potential issues with discriminatory power. Convergent validity was examined using relationship anxiety, trauma symptoms (intrusive thoughts, avoidance, hyper-arousal) and depressive symptomatology, with Jouriles et al. (2009) hypothesizing that verbal/emotional abuse would positively correlate with each of these psychological distress indicators, even when accounting for physical aggression. However, in multivariate analyses controlling for age and physical aggression, verbal/emotional abuse reported on the CADRI and RVI-C was significantly and positively associated only with relationship anxiety, and not with trauma or depressive symptoms.[14] In these models, there was no significant main effect or interaction with sex, and no significant main effect of length of relationship; results were also similar when only considering adolescents with intact relationships, though an association between verbal/emotional abuse and depressive symptoms emerged in this sub-sample.

Together, these two studies highlight the importance of recall period when considering CADRI (and likely other measure) sensitivity: in both studies, more individuals were identified as recipients of aggression using the cumulative method, and Jouriles et al. (2005) found that even when using a one-time, two-month recall, it was more recent aggression that was likely to be reported. Jouriles et al. (2005) also investigated several other variables that may influence

[14] As in Jouriles et al. (2005), RVI-C physical aggression was significantly associated with relationship anxiety and trauma symptoms, but this association was no longer significant when controlling for verbal/emotional abuse and age, as hypothesized by Jouriles et al. (2009).

reliability of reporting, including relationship status and participant-interviewer sex concordance, finding that while most results remained the same when considering these moderating variables, there were some differences; for example, threatening behavior was more likely to be reported to a sex-concordant interviewer at the first assessment period. Thus, considering these (and other) moderating variables is important for future work.

In sum, the CADRI is the most tested of all the scales identified in this review,[15] and has also been subject to the most rigorous investigations. However, even with this body of evidence, questions remain about the reliability of several CADRI sub-scales (threatening behavior, relational aggression and sexual abuse), as well as the most appropriate recall period. Validity evidence for the CADRI generally met stated hypotheses, although evidence about construct validity has been restricted to a limited set of convergent variables, and no study thus far has demonstrated discriminant validity.

Other Measures.

VIolence faite aux Filles dans les Fréquentations à l'Adolescence (VIFFA). Lavoie and Vézina (2001) present a two-part dating violence assessment for adolescents aged 14-16, the VIFFA, which measures received violence for females, and perpetrated violence for males (Table 3.1). Item development was guided by past scales of intimate partner and dating violence, as well as by the recommendations of adolescent focus groups. The first part of the VIFFA is a screen that helps participants identify their most difficult past year dating relationship, in order to facilitate recall (e.g., asking girls if they had ever felt demeaned or treated like an object, and asking boys if they'd ever bullied a girlfriend to get her to do what he wanted; the female screen has five items, and the male screen has eight items). Thinking about this most difficult

[15] Schiff and Zeira (2005) also explored the CADRI in a sample of Israeli youth. While the focus of this article was not measure testing (and so it is not discussed in detail here), they did demonstrate adequate internal consistency reliability of the translated, Hebrew CADRI (all $\alpha > .70$), and that the CADRI-Hebrew was moderately to strongly correlated with CTS-Hebrew sub-scales (*r* range .35-.74).

relationship, participants then respond to 29 items about physical, psychological and sexual victimization (females) and perpetration (males). In Lavoie and Vézina's (2001) sample (see Table 3.1), endorsement of individual items ranged from 0.53-56.53% for victimization, and from 0.33-45.68% for perpetration, with psychological violence the most commonly perpetrated and received.

Using EFA with standardized items under orthogonal rotation, Lavoie and Vézina (2001) found four victimization sub-scales for girls (verbal/emotional abuse, physical abuse, jealousy, and sexual abuse) and four perpetration sub-scales for boys (psychological abuse, jealousy and sexual abuse, severe physical abuse, and minor physical abuse). For both males and females, internal consistency reliabilities on all sub-scales were adequate (range, α, female: .72-.88; range, α, male: .71-.82). By summing the z-scores for each factor, and considering positive sum scores indicative of abuse (i.e., indicating more violence than same-sex peers in the sample), Lavoie and Vézina (2001) found that 28.1% of girls in their sample experienced violence victimization, and 26.8% of boys perpetrated violence.

Discriminant validity analyses were performed using Paulhus's (1984) social desirability scale, assessing self-deception and impression management, in a sub-sample of study participants (n=144); discriminant validity analyses were not performed on the empirically identified sub-scales, but rather on the three intended, theoretical sub-scales (physical, sexual and psychological violence). In this sub-sample, there were no significant associations between either component of social desirability and violence experienced or inflicted, and all associations were small in nature (all $r \leq |0.22|$). In a note at the end of the article, Lavoie and Vézina (2001) also describe a criterion validity assessment, where they conducted interviews with a small sample of boys and girls, in order to determine if responses during 6-month follow-up interviews reflected participants' baseline responses on the VIFFA. Correlations for girls' experience of physical and

sexual violence and boys' use of physical violence were strong ($r \geq 0.75$), but were smaller in magnitude for use (girls) and receipt (boys) of psychological violence ($r < 0.50$). The correlation for boys' use of sexual violence could not be calculated because none of the interviewed boys mentioned sexual violence.

In sum, this study provides good preliminary evidence for the psychometric properties of the VIFFA, including scores calculated using within-sample norms (i.e., z-scores) and crude discriminatory power. Discriminant validity analyses also demonstrated that the VIFFA and social desirability were assessing different constructs, though more validity testing is needed. For broader application, the measure also should be translated to English, and tested in an English-speaking sample. The authors also suggest that several items need revision, and so it is possible that the items presented in the paper are not the final items. Finally, it is not clear whether the empirical or theoretical sub-scales should be used when scoring this scale.

Adolescent Partner Aggression Scale (APAS). Leisen (1999) developed a victimization and perpetration inventory for ethnic minority youth, supplemented with questions about how the actions made the participant feel (e.g., afraid, helpless) and if there were any injuries. Items were developed by reviewing the partner violence literature and previous scales, and also via a thorough content validation procedure. The reading level of this measure is 3.6, and it takes 10-15 minutes to complete. To test this measure, Leisen (2000) conducted several studies with two adolescent samples in Hawaii and Massachusetts (Table 3.1).

In the first study (Table 3.1), Leisen (2000) performed a multi-step content validation, including semi-structured interviews with 25 adolescents who had been in an aggressive relationship, and expert review with ten experts (e.g., clinicians who worked with adolescents, researchers who studied adolescent aggression), who gave both quantitative (e.g., ratings of item wording) and qualitative feedback. The second study was conducted with Hawaiian youth, and

was used to determine final scale content, as well as internal consistency reliability. Using EFA under oblique rotation, Leisen (2000) found a four-factor solution for victimization (mild aggression, control tactics, sexual aggression, and severe physical aggression) and a two-factor solution for perpetration (mild aggression and severe aggression); these final scales contained 32 victimization and 32 perpetration items. Internal consistency for all sub-scales was good (range, α, victimization: .83-.93; range, α, perpetration: .88-.90), though high values of Cronbach's alpha for some sub-scales suggest a narrow scope of item content (e.g., the 9-item severe victimization scale, α=.93). Sub-scale means by sex and ethnicity indicated preliminary discriminatory power, although given the scale range, also indicated that score distributions were likely positively skewed.

Leisen (2000) used Study 3 data to determine internal consistency reliability in a more diverse sample of Massachusetts high school students, to further explore the APAS structure (using EFA; sample size limits precluded CFA), and to obtain validity information. Convergent validity was assessed using the Conflict Tactics Scale (Straus, 1979), and specific hypotheses made about the magnitude of correlations (e.g., that the CTS verbal aggression sub-scale would correlate strongly (*r*=.50-.70) with APAS mild aggression, and moderately (*r*=.30-.50) with APAS severe aggression and sexual aggression; p. 42); however, no hypotheses were made about correlations between the CTS and the APAS control tactics sub-scale. Discriminant validity was assessed using the Marlowe-Crowne Social Desirability scale (Crowne & Marlowe, 1960), with the hypothesis that social desirability would not account for variance in APAS sub-scale scores (p. 43).

In the Study 3 sample, Leisen (1999) found a three-factor structure for victimization (control tactics, physical aggression and sexual aggression), and a two-factor structure for perpetration (mild aggression and severe aggression); thus, the victimization solution differed

from that found in Study 2, with a sample of Hawaiian adolescents. However, correlation analysis provided support for a high degree of congruence between the victimization and perpetration sub-scales found in Study 2 and Study 3. Internal consistency reliability remained adequate (victimization, α, range: .72-.87; perpetration, α, range: .87-.91) and means in this sample also indicated preliminary discriminatory power, though score distributions again appeared highly positively skewed.

All APAS victimization sub-scales (whether calculated using the Study 2 or Study 3 factor solution) correlated strongly with CTS verbal aggression (all $r > .50$), partially supporting hypotheses. The physical and sexual APAS victimization sub-scales (whether calculated using the Study 2 or Study 3 factor solution) also all correlated strongly with CTS physical aggression (all $r > .47$), again partially supporting hypotheses. For APAS perpetration (whether calculated using the Study 2 or Study 3 factor solution), associations were as hypothesized with CTS verbal aggression (stronger association with mild than severe APAS perpetration), but were the opposite of hypothesized associations with CTS physical aggression (stronger association between APAS mild aggression and CTS physical aggression than between APAS severe aggression and CTS physical aggression). When examining associations with social desirability, all correlations were small in magnitude, though associations with the Study 3 control tactics victimization sub-scale (r=-.14, p=.04) and Study 2 and Study 3 mild aggression perpetration sub-scale (r=-.19, $p < .01$ for both) were significant, such that hypotheses were not fully supported. However, preliminary evidence for the reliability and validity of this measure is promising. Future testing should clarify the nature of the scale structure using CFA, and also demonstrate the invariance of this structure by age and gender. Also, since most of the APAS sub-scales correlated more strongly with CTS verbal aggression than CTS physical aggression, item content may need additional review/revision (to ensure the full range of physically

aggressive actions is covered). Finally, it is not entirely clear which 32 items comprise the final scale, and so this should be clarified (the Appendix with the final scale lists 35 items, not 32).

Warning Sign Behaviors (WSBs) / Tendency to Resist or End Abusive Dynamics (TREAD). Murphy, Smith and Xenos (2012) present an inventory of 23 behaviors they call Warning Sign Behaviors (WSBs); these behaviors are based on a dyadic slippery slope model of chronic abuse, where WSBs "can evolve over time and lead to serious harm" (p. 346; see also Murphy & Smith, 2010). For example, the item "he/she picks fights with people they think are trying to steal you" is considered a possessiveness WSB. The goal of the WSB/TREAD scale is not only to detect whether participants have experienced these WSBs, but also to assess what the participant would do if a partner used a particular WSB, information Murphy et al. (2012) feel is useful to prevention and intervention (e.g., for assessing program effectiveness). This latter assessment is referred to as the Tendency to Resist or End Abusive Dynamics (TREAD), and takes 10-20 minutes to complete (1 to 2 minutes per WSB); Murphy et al. (2012) found no evidence of undue participant burden (e.g., no participant expressed discomfort). The TREAD score is assigned by raters, with higher scores indicating more assertive/protective responses (detailed scoring guide is available in Murphy et al., 2012).

The 23 WSBs were developed using focus groups with Australian youth in grades 8-12 (~300 participants), as well as by administering the scale to an online sample of 426 young adult women (aged 18 to 25), in order to ensure all WSBs had been experienced by at least some respondents, and were capable of causing harm. For each WSB, Australian participants in the final adolescent testing sample (n=146, mean age=14.7 years) were asked to write down what they would do if a partner used a WSB, as well as to indicate which WSBs they had experienced.

The TREAD was first presented in Murphy & Smith (2010), where open-ended responses were scored using a 3-point scale. Murphy et al. (2012) refined the nature and scoring of the

open-ended response scale, and also assessed the reliability and validity of this measure. To develop the TREAD scoring rubric, 10 participants' responses were reviewed, to determine typical responses to WSBs; from these 10 participants, five levels of assertiveness were detected (1=*response likely to fuel hostility/response lacks objection*; 5=*response is highly assertive/protective*). To establish intra- and inter-rater reliability, the first author rated 30 TREAD responses until all 30 were rated without indecision. All three authors then rated 30 randomly selected TREAD tests, and reliability calculated using a one-way random-effects intra-class correlation (ICC). The one-way ICC was significant (ICC[1,1]=.93, $p < .01$), indicating high inter-rater reliability.

Validity evidence was focused on content validity (i.e., that the included WSBs could lead to harm), as well as construct and criterion validity. Construct validity was investigated by exploring the empirical structure of the TREAD scores; the hypothesized theoretical structure for the WSBs has five sub-scales (dominance, possessiveness, denigration, conflict-control and retaliation), and so it is possible that the authors believed the TREAD scores would also have five sub-scales, though no specific hypotheses were provided. Criterion validity was assessed by investigating the relationship between the participants' WSB and TREAD scores.

In the online (young adult) sample, prevalence of WSBs ranged from 25% to 69%; participants were also asked about their own use of the 23 WSBs, with prevalence ranging from 5% to 70%. When asked how harmed they felt by the experience of each WSB, the percentage of participants feeling harm ranged from 38% (for "tries to be with you all the time, in time you want to yourself") to 91% (for "hurts your feelings (with words) because you insult, hurt or humiliate them"), with a median harmfulness rating of 83%. Thus, it appears the included WSBs were considered harmful; however, prevalence and perceived harmfulness of these 23 WSBs was not reported for the adolescent testing sample, and so it is not clear whether these results can be

extended to individuals in the target age group.

In the adolescent testing sample, TREAD scores were normally distributed, indicating good discriminatory power. Using this sample, the TREAD scores for the 23 WSBs were submitted to principal components analysis (PCA) and EFA under varimax rotation. These analyses resulted in the removal of four items, with a final three-factor structure (conflict-retaliation TREAD, denigration TREAD and dominance-possessiveness TREAD). Cronbach's alpha for the total 19-item scale was .77, indicating adequate reliability. The three factors were moderately inter-correlated.

As hypothesized, Murphy et al. (2012) found that TREAD scores on each of the three factors, as well as the total TREAD score, were moderately and negatively correlated with recent exposure to WSBs (i.e., more assertive responses to WSBs were associated with the experience of fewer WSBs; *r* range -.29 to -.42). The correlation with total TREAD was stronger when only considering the 66 participants who had experienced WSBs in the past three months; for these 66 participants, as the TREAD score increased, the mean WSB score linearly decreased.

This scale is promising for use as part of program evaluation, but requires further testing. Specifically, the prevalence and perceived harmfulness of the WSBs in an adolescent sample should be explored; more broadly, future research should focus on supporting the assumption that the slippery slope model pertains to teen dating relationships. Also, research about whether higher TREAD scores correspond to increasingly desired and appropriate responses in different cultural contexts, as well as research on the TREAD that includes males, would be valuable.

Recent Fearful Dating Experiences (RFDE). Also examining pre-cursors to dating violence, Schultz and Jaycox (2008) tested an 11-item scale that assesses recent fearful and aversive dating experiences (RFDE), such as "your date made you feel unsafe"; items were adapted from the Women's Experience with Battering Scale (Smith, Earp, & DeVellis, 1995).

Scoring for the 11 items is on a 6-point Likert-type scale (Table 3.1), but dichotomous scores can be created by classifying participants who agree with an item as a 'yes' response. In their representative sample of Latino 9th graders, 40% percent of respondents endorsed at least one RFDE item, with the prevalence of endorsement decreasing as frequency of fearful/aversive experiences increased (e.g., only 5% endorsed ≥ 7 items). There was no difference in total number of fearful experiences by respondent sex, but were some differences on individual items. Specifically, girls were more likely to endorse the two sexual items (e.g., "you were afraid your date might force you into sexual relations that you didn't want"), while boys were more likely to endorse the item "you tried not to 'rock the boat' or cause any trouble because you were afraid of what your date might do." Overall, boys reported more fear across all items than girls (M=1.71 vs. M=1.58, p=.04, respectively); these mean scores also indicate a positively skewed distribution, potentially limiting discriminatory power. The internal consistency of the 11-item scale was .93, which, given the short nature of the RFDE, indicates a narrow item focus that may not fully capture the latent fear construct. Five-day test-retest reliability showed moderate to strong correlations for all items (r range .45 - .60), and a strong correlation on the total score (r = .66), indicating good stability of the total score over time. Where there was disagreement, boys were more likely to change their responses at the re-test; specifically, compared to girls, boys expressed less fear on six of the 11 items at re-test, and so the scale may be less stable for male participants. Reasons for this differential stability should be explored in future research.

To further assess reliability, Schultz and Jaycox (2008) also performed cognitive interviews with a subgroup of participants (n=19, 52.6% female), in order to confirm that these participants understood the meaning of the 11 RFDE items, including wording and key words. These interviews demonstrated that the meaning of the questions was understood by the majority of participants (84%), and that most were able to define the key words (72%). In addition,

approximately half (53%) were able to demonstrate their understanding of the item by describing a concrete situation that could lead to the feeling captured by the item. Finally, these interviews showed that male and female participants interpreted and responded to items in a similar way.

Construct validity was assessed using the attitudes scales presented by Orlando et al. (2006; see Part 2), as well as by asking participants what they would do after hearing a hypothetical scenario depicting psychological and physical aggression (e.g., "I would try to figure out some way to make sure it didn't happen again"; "I would stop dating that person"). Schultz and Jaycox (2008) used these variables to determine which, if any, were associated with being more fearful, but did not present any specific hypotheses. In separate sex-stratified logistic regression models, Schultz and Jaycox (2008) found that girls who were more accepting of male-to-female aggression were more likely to report fearful/aversive dating experiences. For the hypothetical situations, boys who said they would find it hard to stop thinking about the aggression were more likely to report fearful/aversive dating experiences, and both boys and girls who reported they would try not to think about it or pretend it never happened were more likely to report fearful/aversive dating experiences, as were girls who reported they would stop dating that person. No other associations with attitudes or hypothetical behaviors were found. However, since no *a priori* hypotheses were presented, it is not clear whether this pattern of associations was as expected, or if it supports construct validity.

In sum, strengths of this measure include the use of cognitive interviews to confirm that participants understood item content, as well as good reliability. Future research could include more rigorous validity testing, confirmation of a one-factor structure, and testing in more diverse groups (e.g., broader age range, more diverse racial/ethnic background). Since the authors suggest that this measure could be used to assess pre-cursors to dating violence, and since dating violence begins before the 9th grade (i.e., the age group tested in this sample), it would also be

beneficial to assess the measure with middle school students.

Empirical Work

From all empirical searches, 34 articles were found that used an included behavior measure and met inclusion criteria. Approximately 68% of these articles (*n*=23) used some version of the CTS, with a minority using the CADRI (26.5%; *n*=9). The other two articles used the RFDE (Ulloa, Jaycox, Marshall, & Collins, 2004) and the VIFFA (Gagné, Lavoie, & Hébert, 2005). Of these 34 articles, 35.3% (*n*=12) reported on male perpetration, male victimization, female perpetration and female victimization, while a smaller number reported on male and female victimization (23.5%; *n*=8), male and female perpetration (17.6%; *n*=6), and female victimization alone (17.6%; *n*=6). Of the remaining two articles, one reported only on male perpetration, and the other reported only on female perpetration. Information on these empirical articles, including prevalence rates, is presented in Appendices 3.B-3.E.

General observations across these 34 articles include the non-representative nature of most samples (76.5% of articles), the repeated use of several large samples, and the use of modified measures. The one article using the RFDE (Ulloa et al., 2004) drew participants from the same sample that was used to test this measure (and was also the sample used to test several attitudes measures; see Part 2). Articles using the CTS did not typically use the versions that were tested in adolescent populations (i.e., M-CTS and CTS2-Physical Perpetration), and of the articles that did use the tested CTS scales, a majority presented information on the same sample that was used to test these measures (e.g., see Appendix 3.C, footnotes g and r). Articles that used the CADRI used the Spanish and English versions, as well as a Hebrew version, an Arabic version and a Thai version, and one article used a modified CADRI with only two of the original items (Walton et al., 2009). While the study using the VIFFA (Gagné et al., 2005) used a different set of participants than the sample used to test this measure (Lavoie & Vézina, 2001),

the participants in Gagné et al. (2005) were drawn from the same geographic area as the participants in Lavoie and Vézina (2001), and both studies were conducted by the same group of researchers.

Prevalence ranges from these 34 articles are presented in Table 3.2. For both the CTS and CADRI, ranges for physical and psychological victimization and perpetration were quite wide, although this was especially true for psychological victimization and perpetration (Table 3.2). For male victimization, male perpetration, female victimization and female perpetration, median prevalence estimates obtained for physical aggression were similar on the CTS and CADRI, which is not surprising given the overlapping physical aggression items on these two scales. However, median psychological aggression prevalence found on the CADRI was consistently lower than median psychological aggression prevalence found on the CTS, though this difference was most pronounced for victimization estimates (Table 3.2). This is somewhat surprising since the CADRI contains a broader range of psychological aggression items than the CTS; however, it is important to point out that there were a small number of studies that assessed psychological aggression (e.g., for male victimization, three used the CADRI and five used the CTS, and two of the CTS studies drew data from the same sample), and so it is difficult to make definitive statements about differences between the CTS and CADRI based on these articles. For both, however, median psychological aggression prevalence rates were high (range, 72.0-92.0% and 46.7-80.0%, respectively), suggesting that neither possess good discriminatory power (though this depends on what the measures are aiming to assess when they study 'psychological aggression'; see Exner-Cortens, Eckenrode, Schrader, & Rothman, Chapter Five). Very few studies used the CTS or CADRI to assess sexual victimization or perpetration, and so summary statements about differences between these two measures for this behavior are premature; also, the sexual aggression sub-scale of the CTS has not been tested in an adolescent population.

While the median CADRI prevalence rate for female sexual victimization was quite high (30.7%), three of the samples that contributed to this median estimate were from Israel and Thailand, suggesting potential cultural rather than measurement differences. Also, the high estimate for sexual violence victimization on the CTS (70.2%) came from a sample of girls involved in child protective services, again suggesting important sample, and not measurement, differences.

Table 3.2
Summary of Prevalence Estimates: Behaviors

		Range					
	Number of studies	***Low***		***High***		***Median***	
		CTS	*CADRI*	*CTS*	*CADRI*	*CTS*	*CADRI*
Male, victimization							
Psychological	21	11.0	12.2	92.3	94.5	72.0	49.2
Physical		1.9	18.0	41.0	41.9	29.3	30.6
Sexual		-	12.2	-	46.4	-	38.6
Female, victimization							
Psychological	27	4.2	5.8	93.7	95.5	87.9	46.7
Physical		2.7	9.6	67.2	41.2	24.8	18.0
Sexual		2.0	1.9	70.2	46.7	6.6	30.7
Male, perpetration							
Psychological	20	12.1	10.2	92.8	95.3	84.0	80.0
Physical		0.9	4.9	47.6	26.5	17.6	13.6
Sexual		-	-	-	-	-	-
Female, perpetration							
Psychological	20	2.7	3.8	95.3	97.0	92.0	79.1
Physical		2.4	11.3	59.0	41.0	28.6	20.2
Sexual		-	-	-	-	-	-

Note: To be included in summary estimates, more than two studies needed to report on prevalence for the given behavior type/scale combination. The number of studies using probability-based samples was too small to make comparisons by non-representative (n=26) vs. representative (n=8) samples.

Several broad conclusions can be drawn from these empirical articles. The first is that we found only 34 articles published in the past 10 years that met inclusion criteria, which is certainly far fewer than the number published in this period (Foshee & Reyes, 2011). A number of articles were discarded because they did not report prevalence stratified by sex, or because they didn't report prevalence at all; given continuing questions about dating violence prevalence, authors are encouraged to included sex-stratified prevalence rates in their articles. Second, a number of

articles were excluded because they did not use a psychometrically-tested measure of dating violence, and of those that did use a measure with some evidence of validity and reliability, measures were often modified. Thus, it seems a goal for the field is the consistent use of tested measures in their original format; if researchers have needs that are not met by current scales, then development and testing of a new scale that better meets these needs is an important area of research. We surely also failed to pick up several articles that used some version of the CTS, either because the article wasn't indexed properly, or because the article did not use and cite a version of the CTS that was tested in an adolescent population. Moving beyond the measures themselves, our search identified a number of articles from diverse geographic regions, and prevalence rates from these articles suggested potentially important differences in global dating violence prevalence. However, because these articles did not use a consistent scale, it is difficult to determine whether these differences are measure- or sample-driven; improved understanding of cross-cultural differences is an important area for future research. Finally, a number of articles drew participants from one of several large samples, limiting the scope of knowledge about teen dating violence. While these large studies are important, studies in more diverse populations that use tested teen dating violence measures are needed.

Summary: Part 1

In our searches, six dating violence behavior measures with evidence of reliability and validity were located: the CTS (M-CTS and CTS2-Physical Perpetration), the CADRI, the VIFFA, the APAS, WSB/TREAD and the RFDE. Reviewing item content for the English-language measures, we find substantial overlap for the physical and sexual aggression sub-scales, but less overlap for the psychological aggression sub-scales; this latter finding likely reflects the troubled state of psychological aggression measurement (see Exner-Cortens, Eckenrode, Schrader, & Rothman, Chapter Five). Empirical articles primarily used either the CTS or

CADRI, reflecting findings from Capaldi et al. (2012), but these scales were often used in a modified, or in the case of the CTS, non-tested format. Additional discussion of the implications of this review for dating violence measurement can be found in Part 2.

References

See Part 2.

CHAPTER FOUR

MEASUREMENT OF TEEN DATING VIOLENCE: A COMPREHENSIVE REVIEW, PART 2

Measurement of Teen Dating Violence: A Comprehensive Review, Part 2

Deinera Exner-Cortens

Lydia Gill

John Eckenrode

Author Note

Deinera Exner-Cortens, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

Lydia Gill, Department of Human Development, Cornell University.

John Eckenrode, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

This research was supported in part by a Doctoral Foreign Study Award from the Canadian Institutes of Health Research.

Correspondence concerning this article should be addressed to Deinera Exner-Cortens,

CAMH Centre for Prevention Science, 100 Collip Circle, Suite 100, London, ON N6G-4X8.

Contact: dme56@cornell.edu

Abstract

Part 2 of this review presents seven measures that assess attitudes toward teen dating violence, as well as empirical literature that uses these measures. We conclude our review by discussing the implications of these two review papers for teen dating violence measurement and by offering a brief review of promising measures for future study.

Measurement of Teen Dating Violence: A Comprehensive Review, Part 2

As described in Part 1 of this review, teen dating violence research over the past 30 years has primarily focused on understanding prevalence, correlates and outcomes, with less attention paid to measurement of this construct. However, all past findings rest on quality measurement, and so it is important to review the state of teen dating violence measurement in order to understand the work that has occurred and identify the work that needs to be done. Part 1 of this review identified six behavioral measures that have been the focus of psychometric testing; in Part 2, we will review seven attitude measures, along with six empirical studies that use selected attitude measures.

Method

For a full description of the Methods used in this review, please see Part 1. The eight articles discussing an attitude scale were all located during computer-based searches, and present seven unique attitude measures. Information on samples and measures is given in Table 4.1, and information on the type(s) of reliability and non-face validity evidence presented is found in Appendix 4.A. We located six empirical articles that used one of the selected attitude measures, all during computer-based searches. Information from the empirical articles is summarized in Tables 4.2 and 4.3.

Table 4.1
Summary of Measures: Attitudes

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Items and scoring
Attitudes Towards Dating Violence						
Hokoda et al. (2006)	Attitudes Towards Male Dating Violence – Spanish [AMDV] and Attitudes Towards Female Dating Violence – Spanish [AFDV]	307	62.5[a]	Mexican	Mexicali and Monterrey, Mexico	Number of items: n/a Number of sub-scales: 6 Response options: 5-point Likert-type scale (1=*strongly disagree*, 5=*strongly agree*)
Price et al. (1999)	Attitudes Towards Male Dating Violence [AMDV] and Attitudes Towards Female Dating Violence [AFDV]	823	46.0	Primarily Caucasian (84.0% English-Canadian, 6% French-Canadian or Acadian)	New Brunswick, Canada	Number of items: 76 Number of sub-scales: 6 Response options: 5-point Likert-type scale (1=*strongly disagree*, 5=*strongly agree*)
AADS/JVCT						
Muñoz-Rivas et al. (2011)	Attitudes About Aggression in Dating Situations – Spanish [AADS] and Justification of Verbal/Coercive Tactics Scale – Spanish [JVCT]	2,856	54.8	Spanish	Madrid, Spain	Number of items: AADS – 10; JVCT – 24 Number of sub-scales: AADS – 3; JVCT – 6 Response options: AADS – 6-point Likert-type scale (1=*strongly agree*, 6=*strongly disagree*); JVCT – 5-point Likert-type scale (1=*not justified no matter what*, 5=*justified in many situations*)

(Table 4.1 continues)

(Table 4.1 continued)

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Items and scoring
AADS/JVCT						
Slep et al. (2001)	Attitudes About Aggression in Dating Situations [AADS] and Justification of Verbal/Coercive Tactics Scale [JVCT]	2,188[a,b]	49.8[a]	Multi-racial (56.2% Caucasian, 17.3% Hispanic, 13.6% African American)[a]	Suffolk County, NY	Number of items: AADS – 10; JVCT – 24 Number of sub-scales: AADS – 3; JVCT – 6 Response options: AADS – 6-point Likert-type scale (1=*strongly agree*, 6=*strongly disagree*); JVCT – 5-point Likert-type scale (1=*justified in MANY situations*, 5=*not justified NO MATTER WHAT*)
Other – Attitudes						
Davidson (2005)	Adolescent Attitudes Regarding Dating Relationships [AARDR]	St 1: 18 St 2: 600[c] St 3: 600[c] St 4: 125	St 1: 50.0 St 2 and St 3: 50.0 St 4: 68.8[a]	St 1: Multi-racial (44.4% multi-racial or other, 27.8% White, 27.8% Black)[a] St 2 and St 3: Primarily Caucasian (90%) St 4: Primarily Caucasian (88%)	Columbia and Jefferson City, MO	Number of items: Male – 36; Female – 27 Number of sub-scales: Male – 5; Female – 3 Response options: 6-point Likert-type scale (1=*strongly agree*, 6=*strongly disagree*)
Edelen et al. (2009) and Orlando et al. (2006)	Prescribed Norms Scale and Approval of Retaliation scale [NOBAGS]	2,575[d]	51.0[a]	91% Latino	Los Angeles, CA	Number of items: 9[e] Number of sub-scales: 2 Response options: 4-point Likert-type scale (1=*strongly disagree*, 4=*strongly agree*)[f]

(Table 4.1 continues)

(Table 4.1 continued)

Authors and year	Measure name [Abbreviation(s)]	N	% female	Racial/ethnic distribution	Location	Items and scoring
Other – Attitudes						
Rayburn et al. (2007)	Articulated Thoughts in Simulated Situations [ATSS]	41[d]	36.6[a]	98.5% Latino	Los Angeles, CA	Number of items: 4 scenarios, 5 segments Number of sub-scales: n/a Response options: Open-ended
Slep et al. (2001)	Attitudes Towards Interpersonal Violence [AIV]	2,188[a,b]	49.8[a]	Multi-racial (56.2% Caucasian, 17.3% Hispanic, 13.6% African American)[a]	Suffolk County, NY	Number of items: 6 Number of sub-scales: 2 Response options: 5-point Likert-type scale (1=*never*, 5=*always*)

Abbreviations: St – Study

[a]Hand calculated

[b]This sample was also used in Cascardi et al. (1999) [see Part 1]. Per Slep et al. (2001), the Cascardi et al. (1999) sample contained only current and recently dating females and currently dating males, while Slep et al. (2001) also included recently dating males. Also, participants were included in Cascardi et al. (1999) if they were missing the AADS, JVCT or both, but these participants were excluded from Slep et al. (2001).

[c]Randomly selected sub-sample of a larger sample. Original sample had 595 male and 802 female participants, and Davidson (2005) randomly selected 300 males and 300 females from this larger sample for St 2 and St 3. The remaining individuals were preserved for use in confirmatory factor analysis (data not reported).

[d]Participants were drawn from the same sample. This sample was also used in Schultz & Jaycox (2008) [see Part 1].

[e]In Orlando et al. (2006), scale contained 14 items.

[f]These are the response options for the Prescribed Norms Scale. For NOBAGS, the response options were 1=*really wrong* to 4=*perfectly OK*.

Results

Measures

Attitudes Towards Dating Violence Scales (ATDVS). Price, Byers et al. (1999) developed the Attitudes Towards Male Dating Violence (AMDV) and Attitudes Towards Female Dating Violence (AFDV) scales to assess attitudes toward heterosexual physical, psychological, and sexual dating violence among adolescents of all ages. Scale items were generated using existing scales, by reviewing the literature, and by adolescents, teachers, and members of the research team. The initial scales were reviewed by adolescents and teachers, piloted on ten 14-19 year olds, and then further developed using two test samples. The first test sample (*n*=158 Francophone and *n*=350 Anglophone) was comprised of university students, and was used to determine internal consistency reliability and factor structure, and to refine item content. The second test sample was comprised of 211 Francophone and 154 Anglophone high school students, and was used to finalize scale content. Items are written in both French and English, but only information on the English scales was presented in Price et al. (1999). Although the AFDV and AMDV scales contain some similar items, they were developed separately to account for differences in how violence might be used by boys and girls.

Using a non-representative sample of 823 grade 7, 9 and 11 students from New Brunswick (Table 4.1), Price et al. (1999) found that each of the six theoretical ATDV scales had a one-factor structure under principal axis factoring (this structure was tested by randomly splitting the sample in half, and checking the structure in each half of the sample. Cattell's salient similarity index indicated that the factor loadings were highly similar in both groups, and so they were combined in subsequent analyses). The six scales are called AMDV-Phys (attitudes toward male use of physical dating violence; e.g., "Some girls deserve to be slapped by their boyfriends"; 12 items, α=.83), AMDV-Psyc (attitudes toward male use of psychological dating

violence; e.g., “A girl should always do what her boyfriend tells her to do”; 15 items, α=.83), AMDV-Sex (attitudes toward male use of sexual dating violence; e.g., “It is alright to pressure a girl to have sex if she has had sex in the past”; 12 items, α=.87), AFDV-Phys (attitudes toward female use of physical dating violence; e.g., “Some guys deserve to be slapped by their girlfriends”; 12 items, α=.85), AFDV-Psyc (attitudes toward female use of psychological dating violence; e.g., “Girls have a right to tell their boyfriends what to do”; 13 items, α=.75) and AFDV-Sex (attitudes toward female use of sexual dating violence; e.g., “To prove his love, it is important for a guy to have sex with his girlfriend”; 12 items, α not reported). In the full sample, two items had factor loadings below 0.30, one on the AFDV-Psyc scale (“A girl should not control what her boyfriend wears”) and one on the AFDV-Phys scale (“Pulling hair is a good way for a girl to get back at her boyfriend”), but were kept because of their stronger loadings on the French version of the scales. Internal consistency reliabilities for the total sample, and by sex, were adequate for each of the six scales (α, by sex, range: .72-.88). On four of the six scales (all except AFDV-Psyc and AFDV-Phys), internal consistency reliabilities were slightly higher for males than females.

For ATDVS construct validity, Price et al. (1999) hypothesized that mean scores on the AFDV scales would be higher than the scores on the corresponding AMDV scales for both boys and girls, but that boys would be more accepting of violence overall than girls. Construct validity was also assessed by examining the inter-correlations between the ATDV scales, as well as their correlation with a measure of gender role attitudes (the Attitudes Towards Women Scale for Adolescents; Galambos, Peterson, Richards, & Gitelson, 1985), predicting that males and females who held more traditional gender role attitudes would be more accepting of boys’ and girls’ use of dating violence, and that the AMDV sub-scales would be positively associated with each other, as would the AFDV sub-scales. Criterion validity was assessed by examining own

and friend use of dating violence, with the hypothesis that participants who were more accepting of dating violence would use more physical, sexual and psychological dating violence in their relationships (with specific hypotheses for each AMDV and AFDV scale; pp. 354-355), and would also have friends who used more physical and sexual dating violence. Mean scores on each of the scales demonstrated preliminary discriminatory power (i.e., distributions appeared approximately normally distributed).

Price et al. (1999) found that older students were less accepting of dating violence than younger students, although they did not make any specific hypotheses about age effects. For attitudes toward male dating violence, participants were most accepting of AMDV-Psyc, followed by AMDV-Phys and AMDV-Sex, while for attitudes toward female dating violence, participants were most accepting of AFDV-Phys, followed by AFDV-Psych and AFDV-Sex. As hypothesized, male and female participants were more accepting of female than male psychological, physical and sexual violence, and male participants also had significantly higher (i.e., more accepting) scores on all six scales compared to female participants. Price et al. (1999) also found that greater acceptance of dating violence on five of the six ATDV scales was moderately to strongly associated with more traditional gender role attitudes for both male and female participants (r range |.34|-|.59|), with the correlation between AFDV-Phys and gender role attitudes slightly smaller in magnitude (r=|.23|). The six scales were also significantly inter-correlated in both the male and female sub-samples, with most correlations in the moderate to strong range, though the correlations between AFDV-Phys and AMDV-Sex, and AFDV-Phys and AFDV-Sex, were smaller in magnitude for male participants (r=.24 and r=.24, respectively).

Criterion validity hypotheses were partially supported. Associations between boys' attitudes toward dating violence and use of dating violence behavior (e.g., between AMDV-Psyc and psychological aggression perpetration) were significant, and small to moderate in magnitude

(*r* range .23-.39 for hypothesized associations), but associations between boys' attitudes and friends' use of dating violence were smaller (*r* range .09-.26), and several failed to reach statistical significance. For girls, most associations between attitudes and use of dating violence were moderate in nature (*r* range .27-.35 for hypothesized associations), but the association between AFDV-Sex and psychologically abusive behavior was smaller (*r*=.11) and failed to reach statistical significance (correlations with sexually abusive behavior could not be calculated for girls due to the small sample size). For associations between girls' attitudes and friends' use of dating violence, most were very small in magnitude (*r* range .01-.18), and none reached statistical significance.

Hokoda et al. (2006) translated the ATDVS into Spanish, and tested its reliability in a sample of Mexican high school adolescents (noting that not all Mexican adolescents go to high school, and so this is a somewhat selective sample). The translated measure was blindly back-translated, and translations were discussed between the translators and shown to Mexican social service professionals who work with youth. The translated measure was also piloted with Mexican adolescents. Less information was given on the ATDVS translation than on the CADRI translation with this same sample (see Part 1); specifically, no information was given about any wording or reading-level changes. Two-week test-retest reliability was strong for five of the six scales in both males and females (*r* range .51-.71), but was weaker for females on the AMDV-Sex scale (*r*=.43). The AMDV scales in the female sub-sample were also the only scales with test-retest reliabilities below .60. Finally, test-retest reliabilities were higher for males and females on sex-concordant attitude scales (e.g., boys had higher test-retest reliability on the AMDV scales, compared to girls). Internal consistency reliabilities for males and females were adequate (> .70) for all sub-scales, except male attitudes toward female-perpetrated psychological abuse (α=.67), and female attitudes toward male-perpetrated sexual (α=.53) and

psychological (α=.57) abuse. Internal consistency reliabilities were higher for males and females on sex-concordant attitude scales, except for AFDV-Sex, which was slightly more reliable for boys than girls. No information was presented on the structure or validity of the translated measure, although the authors state that it is "not clear that the six subscales assess independent constructs" (p. 123), and that there may be comprehension difficulties with the translated scales because of the many colloquial items (e.g., "bad mouth", "get back at"). They also discuss that body language of some participants suggested the measure was difficult to understand and too long (was administered after the CADRI in this study; see Part 1).

These two studies demonstrated adequate reliability of the ATDVS, although scales may be less reliable when completed by a sex-discordant sample (e.g., AMDV and a female sample). Price et al. (1999) provided evidence of construct validity, meeting hypotheses about mean scores and associations with gender role attitudes, but evidence of criterion validity was slightly weaker, especially for girls. Also, given that sex and age affected attitudes, it will be important in future study to demonstrate invariance of the six-scale structure across these groups. Finally, Price et al. (1999) state that their study also aimed to establish normative levels of attitudes, and the establishment of norms in an important goal; however, to meet this goal, this scale should be administered to a representative sample of adolescents.

Attitudes About Aggression in Dating Situations (AADS) and Justification of Verbal/Coercive Tactics Scale (JVCT). Slep, Cascardi, Avery-Leaf and O'Leary (2001) created two new attitude measures that were designed to elicit a broad range of responses (to address floor effects), assess attitudes toward male and female use of physical and psychological aggression, and consider attitudes about aggression against dating partners and peers (because these might be related). The first new scale, AADS, assesses physical aggression, and includes antecedent situational cues (e.g., "Mark calls Tina a slut in front of their friends. Tina slaps

him"), where respondents are asked to indicate their (dis)agreement with the underlined portion of the question. The second new scale, JVCT, assesses attitudes toward use of three types of psychological aggression (verbal aggression, controlling behaviors, and jealous tactics).

AADS items were developed by reviewing the empirical literature and by showing potential items to a female focus group (n ~= 15). The final AADS has 32 questions, resulting from crossing sex of perpetrator (2 levels), type of aggressive action (4 levels), and type of provocation (4 levels). Due to time constraints, the AADS tested in Slep et al. (2001) contains a random selection of 10 of these 32 questions, where 5 represented male aggression and 5 represented female aggression, and each type of behavior and provocation appeared at least once. They also kept two items describing same-sex peer-to-peer aggression, for a total of 12 items. JVCT items were adapted from the Conflict Tactics Scale (verbal aggression items; Straus, 1979), and from the Dominating and Jealous Tactics scale (controlling and jealous behavior items; Kasian & Painter, 1979). The JVCT is comprised of two sets of 12 parallel items assessing male-to-female and female-to-male aggression, respectively, with an equal number of items on verbal aggression, jealous behaviors and controlling/dominating behaviors (i.e., four per scale).[16] Higher scores on the AADS and JVCT indicate less acceptance of the behavior.

Construct validity was assessed by examining the factor structure of the new measures; Slep et al. (2001) hypothesized a three-factor structure for the AADS (AADS-Female Aggression, AADS-Male Aggression and AADS-Peer Aggression) and a six-factor structure for the JVCT (JVCT-Female Verbal Aggression, JVCT-Male Verbal Aggression, JVCT-Female Jealous Tactics, JVCT-Male Jealous Tactics, JVCT-Female Control Tactics and JVCT-Male Control Tactics; see p. 308 for further details). For AADS factor analyses, the full sample was randomly split into two sub-samples, to conduct both EFA (orthogonal rotation) and CFA, with

[16] The Appendix of Slep et al. (2001) only contains 11 JVCT questions, but the Methods section specifies that there are 12 parallel items.

analyses run separately by sex.

As hypothesized, both EFA and CFA supported a three-factor AADS solution for males and females.[17] Internal consistency reliabilities for these three AADS sub-scales were good (range, α, male: .83-.87; range, α, female: .79-.86), and two-week test-retest reliabilities were fairly strong, though were higher on all sub-scales for females than males (range, *r*, male: .57-.63; range, *r*, female: .65-.74),[18] and only the test-retest reliability for females on the AADS-Female Aggression sub-scale exceeded Kline's (2000) recommended cut-off of .70. Skewness for the three AADS sub-scales indicated a slight positive skew for males' and females' attitudes toward female aggression, and a slight negative skew for males' attitudes toward peer aggression (all < |.60|), with a larger negative skew for males' and females' attitudes toward male aggression, and females' attitudes toward peer aggression (> -1.05). While some of these values fell within rough guidelines for moderate skew (where low skew is values between ± 0.50, moderate skew is between |.50|-|1.0| and large skew is > |1.0|; Bulmer, 1979), these guidelines are for population skewness; to interpret sample skewness, a sample test statistic can be calculated using the standard error of skewness (Cramer, 1997). Calculating this test statistic and comparing it to the normal distribution, all skewness values for the AADS were found to exceed the critical value (±1.96), demonstrating that all AADS distributions were significantly skewed. However, all sub-scales did have smaller skewness than scores on the Attitudes Towards Interpersonal Violence (AIV) scale, another measure of attitudes toward physical dating aggression (see below), indicating that AADS scores were more normally distributed than AIV scores and that the AADS had greater discriminatory power than the AIV in this sample,

[17] Two items, one specifying female-to-male aggression ("Michelle gets really angry at Carlos for ignoring her, so she hits him to get his attention") and the other male-to-female aggression ("Jenny and Dan are arguing because Jenny wants to see other guys. She gets really mad and starts to hit Dan. Dan grabs Jenny and pushes her away") did not load on any factor, and so were removed prior to CFA analyses, reducing the total number of items to 10 (four on the female aggression sub-scale, four on the male aggression sub-scale and two on the peer aggression sub-scale).
[18] Since these data were from a program evaluation, and the re-test was done at the same time as the program post-test, only data from the control participants (n=668) were used for test-retest calculations.

especially for attitudes toward females' use of aggression.

For the JVCT, CFA was used to confirm the proposed six-factor structure, using the two random sub-samples to assess measurement invariance; for both males and females, this structure showed good fit, with measurement invariance across the two sub-samples. Internal consistency reliabilities were good (range, α, male: .71-.86; range, α, female: .72-.83), and test-retest reliabilities were moderate to strong (*r* range, male: .34-.57; *r* range, female: .41-.69). For all six sub-scales, test-retest reliabilities were stronger for females than males, but none exceeded Kline's (2000) recommended .70 cut-off; stability was lowest for the control tactics sub-scales in both males and females ($r \leq .44$ for JVCT-Female Control Tactics and JVCT-Male Control Tactics; all other JVCT sub-scale test-retest *r*'s were $\geq .50$). Skewness values for four of the six JVCT sub-scales were slightly negative but close to zero for both males and females (range, |.20|-|.55|); however, the sample skewness test statistics for these sub-scales all exceeded the critical value, indicating limited discriminatory power. The skewness values for the two control tactics sub-scales were large and negative for both males and females (range, |2.76|-|3.88|), and the sample test statistics far exceeded the critical value, indicating that these sub-scales had poor discriminatory power.

Convergent validity was assessed by correlating AADS and JVCT scores with the participant's use of physical aggression, and JVCT scores with the participant's use of psychological aggression. Slep et al. (2001) hypothesized that attitudes and behaviors would share 4-16% of variability (i.e., correlations in the range .20-.40), and that associations would be larger within than between sex (e.g., for females, associations between JVCT female sub-scales and behavior were hypothesized to be larger than between JVCT male sub-scales and behavior); however, associations between sex-discordant scales were not presented, and so it does not appear this latter hypothesis was addressed. Slep et al. (2001) also hypothesized that associations

would be larger for concordant attitudes and behaviors (e.g., justification of control tactics and controlling acts) than discordant attitudes and behaviors.

For female participants, associations between the six attitudes scales and behavior were generally in the hypothesized range; however, associations between JVCT-Female Control Tactics and use of physical and verbal aggression were smaller ($r < .10$). Also as hypothesized, associations were largest in magnitude for concordant attitudes and behaviors. Similar results were seen in the male sub-sample, although correlations between JVCT-Male Control Tactics and use of physical and verbal aggression were larger than in the female sub-sample (r=.17 and r=.14 in the male sub-sample, respectively). Thus, construct validity hypotheses were generally supported, though evidence was slightly weaker for the JVCT Control Tactics sub-scales.

Concurrent validity was assessed using another attitude measure, the AIV, as a benchmark (see below for a description of AIV reliability and validity in an adolescent sample). For both males and females on the AADS, sex-concordant correlations (e.g., between AIV-Female Aggression and AADS-Female Aggression for females) were stronger than sex-discordant correlations (e.g., between AIV-Female Aggression and AADS-Male Aggression for females; *r,* sex-concordant, range: .32-.40; *r,* sex-discordant, range: .08-.23). For the JVCT, correlations with the AIV were also small to moderate (r range: .21-.36), but sex concordant vs. discordant differences were not reported.

As with other attitudes scales, physical and psychological female aggression was consistently reported as more acceptable than male aggression in the Slep et al. (2001) sample, and female respondents reported attitudes that were less accepting of dating violence than males on all sub-scales except JVCT-Control Tactics. Because of issues with the stability of the JVCT Control Tactics sub-scales as well as the small magnitude of correlations between these sub-scales and convergent validity measures, however, Slep at al. (2001) concluded that the

controlling tactics items could be omitted without losing any measure usefulness. Also, students in this sample were mostly in grades 11 and 12 (> 90%), and so future research using younger adolescents is warranted. In future work, the 22 omitted AADS questions should also be explored. Finally, the convergent validity measures of psychological aggression were the same as those used to create the JVCT, and so additional evidence of convergent validity for this scale is desirable.

Muñoz-Rivas, Gámez-Guadix, Fernández-González and Lozano (2011) assessed Spanish versions of the AADS and JVCT in a sample of Spanish high school students. Translation of both scales was completed by a group of bilingual researchers, and the translated versions were piloted in a group of 110 students, to ensure comprehension of translated items. Muñoz-Rivas et al. (2011) hypothesized that they would confirm the AADS and JVCT factor structures identified by Slep et al. (2001), and that they would find positive correlations between AADS and JVCT scores and use of physically and psychologically aggressive tactics (as assessed on the M-CTS (Cascardi et al., 1999) and the Dominating and Jealous Tactics scale (Kasian & Painter, 1992), both of which were used previously in a Spanish adolescent and young adult population by Lozano (2009)). As in Slep et al. (2001), Muñoz-Rivas et al. (2011) also hypothesized that they would find stronger correlations between concordant attitudes and behaviors.

Under CFA, Muñoz-Rivas et al. (2011) found adequate fit for the three-factor AADS structure for both males and females. All items had factor loadings ≥ .40, except the loading for the item "David is following Maria and won't leave her alone. Maria pushes him out of her way" from the Female Aggression factor, which was lower in the female sub-sample (λ=.31). Internal consistency reliabilities for males and females on the three AADS sub-scales were all adequate, though some were a bit low, potentially because of the small number of items on each sub-scale (two to four items per sub-scale; range, α, male: .67-.81; range, α, female: .65-.79); internal

consistency reliabilities were higher for males than females on all sub-scales. Muñoz-Rivas et al. (2011) also found good fit for the six-factor JVCT structure for both males and females. All items had factor loadings > .40, except in the female sub-sample for the items "keeping him from doing things to help himself" (λ =.25 on Female Control Tactics) and "keeping her from doing things to help herself" (λ =.28 on Male Control Tactics). Internal consistency reliabilities for the JVCT sub-scales were adequate for the Verbal Aggression and Jealous Tactics sub-scales for both males and females (range, α, male: .71-.80; range, α, females: .66-.75). Internal consistency reliabilities on the Control Tactics sub-scales were lower, and were especially poor for female participants (range, α, male: .60-.69; range, α, female: .34-.39). For all JVCT sub-scales, internal consistency reliabilities were higher for males than females. The low reliabilities for the Control Tactics sub-scales in the female sub-sample, as well as issues with low item loadings, suggests that these sub-scales may need additional work before use with Spanish female adolescents.

As hypothesized, the largest correlations between attitudes and behaviors were found for attitude-behavior concordant pairs (e.g., attitudes toward physical aggression and use of physical aggression); these correlations were generally small to moderate in nature, though the correlation between female attitudes toward jealous tactics and their use of these tactics was stronger (r=.56). The smallest correlations were between female attitudes toward control tactics and use of control tactics and male attitudes toward physical aggression and use of physical aggression (r=.13 and r=.16, respectively). Correlations between attitude-behavior discordant pairs were small to moderate (r range, .05-.29).

Different than the Slep et al. (2001) findings, Muñoz-Rivas et al. (2011) found that males were more accepting than females of physical aggression and controlling behaviors, but that females were more accepting than males of verbal aggression and jealous tactics. Further, males and females were more accepting of physical and verbal aggression when it was used by a girl,

but more accepting of controlling behavior when it was used by a boy (there was no difference for use of jealous tactics). No *a priori* hypotheses were presented for these mean scores, and so it is not clear whether these differences were anticipated.

The Muñoz-Rivas et al. (2011) study demonstrated that these scales are promising for use in Spanish populations, but work is needed on the female JVCT-Control Tactics sub-scales, reflecting findings of Slep et al. (2001). The participants in this study also may have been from highly educated families (as selected high schools needed to teach pre-university studies), and so future work could involve more diverse Spanish populations. Finally, the Spanish versions of the scales weren't presented by Muñoz-Rivas et al. (2011), and mean scores on sub-scales weren't reported, and so translation quality and discriminatory power, respectively, could not be assessed from this article.

While both Slep et al. (2001) and Muñoz-Rivas et al. (2011) provided preliminary support for the reliability and validity of the AADS and the JVCT verbal aggression and jealous behaviors sub-scales, potentially important cultural differences also emerged. For example, the mean score patterns observed in Slep et al. (2001) were somewhat different in Muñoz-Rivas et al. (2011), suggesting cultural differences in attitudes toward teen dating violence. Also, although the scale structures were confirmed, there may still be differences in content interpretation; for example, Slep et al. (2001) observed a moderate correlation between boys' attitudes toward physical aggression and their use of physical aggression, while Muñoz-Rivas et al. (2011) found only a small correlation between these two scores. Thus, the effect of cultural differences on scale responses is an important topic for future research.

Other Measures.

Attitudes Towards Interpersonal Violence (AIV). Although it was not the focus of their paper, Slep et al. (2001) also investigated the psychometric properties of the AIV scale, and so

their findings are discussed briefly here. Like the AADS, the six-item AIV was designed to measure attitudes about the acceptability of physical aggression; however, scores on the AIV are typically very skewed (strong floor effect), while the AADS was designed to mitigate this problem. AIV items are intended to form two sub-scales (AIV-Female Aggression and AIV-Male Aggression), but this structure was not explored in Slep et al. (2001). The internal consistency reliability of the AIV in this sample was good (range, α, male: .79-.82; range, α, female: .82-.83). However, two-week test-retest reliability was weaker than re-test reliability for the AADS (*r*, range males, AIV: .53-.62; *r*, range females, AIV: .43-.58), though, unlike the AADS, the AIV was more stable for males than females. Skewness values for the AIV indicated a large negative skew (range, |1.88|-|5.21|; also supported by sample skewness test statistics) and thus this measure possesses less discriminatory power than the AADS. Like the AADS, however, females were less accepting of violence than males and both males and females were more accepting of male aggression than female aggression. The AIV was moderately correlated with use of physical aggression for both males and females, and the correlation with physical aggression was larger in magnitude than correlations with verbal aggression, jealous behavior or controlling tactics. There were also small to moderate correlations between the AIV and the AADS, and the AIV and the JVCT, indicating these measures do not represent identical constructs.

Slep et al. (2001) also assessed a second-order factor structure using EFA (oblique rotation) for the AIV, AADS and JVCT factor and scale scores, and found a three-factor solution for both males and females (Factor 1-JCVT Female and Male Verbal Aggression and Jealous Tactics; Factor 2-JVCT Female and Male Control Tactics; Factor 3-all AADS and AIV aggression sub-scales). Since all AADS and AIV scores loaded on the same factor, Slep et al. (2001) suggest that researchers can select between the two measures "without sacrificing an

adequate assessment of the construct" (p. 316). As described above, the AADS had greater stability and discriminatory power than the AIV in this sample, but is also a longer measure.

Adolescent Attitudes Regarding Dating Relationships (AARDR). Davidson's (2005) AARDR scales also assesses attitudes toward violence in high school students, with one scale for males' attitudes toward dating violence and healthy dating relationships and another for females' attitudes toward dating violence and healthy dating relationships. Items were created based on focus group data (one male and one female focus group, each with four participants), as well as from a review of prior research and theory, with the goal of including items that represented all levels of the social ecology. Preliminary items were reviewed by a panel of seven experts (e.g., doctoral-level therapist, rape crisis counselor), and pilot testing of the preliminary scale was done with 10 high school students (50% female). Once a final scale was identified (90 items for males and 90 items for females), psychometric testing was performed with a non-representative sample drawn from one public and one private high school (Table 4.1). The reading level of the 90-item scale was 5.9, and it took participants 20-30 minutes to complete the relevant 90-item scale and a demographic questionnaire.

The psychometric testing sample was stratified by sex and 300 participants from each sex-stratified sample were then randomly selected for EFA analyses; the remainder of each sub-sample was reserved for CFA, but these analyses were not reported by Davidson (2005). Reliability of scale responses was assessed by five items that asked participants to report on their truthfulness in completing AARDR items; individuals who "responded inaccurately" to one or more of these items were removed from the sample (p. 70; a comparison between these participants (n=64) and those who did respond accurately was not reported). Individuals in the 600 person sub-sample (300 males, 300 females) completed one of three concurrent validity measures, assessing attitudes toward violence (interpersonal and non-interpersonal domains;

Lonsway & Fitzgerald, 1995), hostility toward women (Lonsway & Fitzgerald, 1995) or rape myth acceptance (Illinois Rape Myth Acceptance Scale-Short Form; Payne, Lonsway, & Fitzgerald, 1999). To assess discriminant validity, participants completed the BIDR-Impression Management scale (Paulhus, 1991). Test-retest reliability was assessed in a sub-sample from one of the high schools (Table 4.1).

Davidson (2005) reports that the AARDR was univariate normal in the total sample (p. 70). In the sub-sample used for EFA, individual items with skewness or kurtosis greater than 2.0 were removed prior to analyses; this resulted in the removal of 22 items for males and 29 items for females. Items were also removed if they had a factor loading of < 0.35, or if they cross-loaded > 0.35. Using principal axis factoring and EFA (orthogonal rotation), Davidson (2005) found a five-factor solution for the remaining 36 items for males; these factors were called Coercive Attitudes (15 items; α=.86), Precursors to Coercion (5 items; α=.76), Peer Pressure (6 items; α=.66), Societal Pressure (5 items; α=.66) and Support for Healthy Attitudes (5 items; α=.58). Sub-scale inter-correlations were all ≤ |0.20|, except for the correlation between Coercive Attitudes and Precursors to Coercion (r = .41) and Peer Pressure (r = .36), which were moderate. Item-total correlations were generally small to moderate, but the correlation between Coercive Attitudes and the total scale was quite large (r = .80), likely because this sub-scale contained 15 of the 36 total items. The two-week test-retest reliability for the total AARDR-Male scale was r =.81, with sub-scale reliabilities ranging from .54 to .82; the lowest test-retest reliability was for the Peer Pressure scale (r=.54), with all other sub-scale reliabilities ≥ .69. The skewness of the final AARDR-Male and most sub-scales was small (≤ |.61|), but the skewness for the Coercive Attitudes sub-scale was larger (-1.06), indicating potential limitations of discriminatory power for this sub-scale. Sample skewness test statistics indicated significant skew in all male total scale and sub-scale scores except Precursors to Coercion.

For females, a three-factor solution was chosen for the remaining 27 items. These factors were called Precursors to Coercion (12 items; α=.83), Peer and Societal Pressure (10 items; α=.75) and Support for Healthy Attitudes (5 items; α=.54); the internal consistency reliability for Support for Healthy attitudes was low, potentially because of the small number of items. Sub-scale inter-correlations were all ≤ |0.21|, and item-total correlations were moderate to strong (*r* range: .40-.69). The two-week test-retest reliability for the total AARDR-Female scale was .78, with sub-scale reliabilities ranging from .77 to .82. Skewness of the final AARDR-Female and all three sub-scales was small (≤ |.39|), but sample skewness test statistics indicated significant skew for Peer and Societal Pressure and Support for Healthy Attitudes scores. For both males and females, higher scores on AARDR sub-scales indicated more positive responses (i.e., less acceptance of dating violence, more awareness of peer and societal pressure, more social support).

All male sub-scales, and the AARDR-Male total score, were weakly to moderately correlated with the BIDR-Impression Management scale (*r* range |.19| to |.37|). For concurrent validity, the total AARDR-Male was most strongly correlated with rape myth acceptance (*r*= -.70), followed by hostility toward women (*r*=-.49), and was weakly correlated with the attitudes toward violence, interpersonal domain scale (*r*=-.27). Correlations between the Coercive Attitudes sub-scale and the concurrent validity measures followed the same pattern as the total scale. Precursors to Coercion and Support for Healthy Attitudes were moderately and negatively correlated with rape myth acceptance, and Precursors to Coercion was also moderately and negatively correlated with hostility toward women and attitudes toward violence (interpersonal domain). All other associations between AARDR male sub-scales and concurrent validity measures were small (*r* ≤ |.24|).

All female AARDR sub-scales, and the total AARDR-Female score, were weakly to

moderately correlated with BIDR-Impression Management (*r* range: |.15|-|.34|). Precursors to Coercion and Peer and Societal Pressure were moderately correlated with rape myth acceptance and hostility toward women, and Precursors to Coercion was also moderately correlated with attitudes toward violence (non-interpersonal domain). All other associations between AARDR female scores (on the total scale and sub-scales) and concurrent validity variables were small in magnitude ($r \leq |.24|$).

This preliminary exploration of the AARDR suggests the scale has promise for use with adolescent samples, following some additional work. For example, although Davidson (2005) states that associations with the BIDR-Impression Management scale provide evidence for discriminant validity, all correlations but one were significant with $p < .01$ (and the other with $p < .05$), and some correlations were moderate in magnitude, suggesting that discriminant validity requires additional evidence. Also, internal consistency reliabilities for some sub-scales were low, though most sub-scale scores were stable over a two-week period. Additionally, although it is intended for high school students (i.e., grades 9-12), the scale was only tested in participants from the 11th and 12th grades, so future testing should involve younger participants. Finally, because no hypotheses were presented regarding expected patterns between AARDR sub-scales and concurrent validity variables, it is difficult to assess this evidence, and the AARDR-Female total scale only weakly correlated with all three concurrent validity variables, suggesting additional work with this scale is needed. Future work could also use existing measures of adolescent attitudes toward dating violence (e.g., the AADS) to assess convergent validity.

Prescribed Norms/NOBAGS. Edelen, McCaffrey, Marshall and Jaycox (2009) and Orlando, Jaycox, McCaffrey and Marshall (2006) used item-response theory to examine items from two existing attitudes scales, the Prescribed Norms scale (Foshee et al., 1996) and the NOBAGS scale (Huesmann & Guerra, 1997), both of which assess approval of aggression

(Table 4.1). Orlando et al. (2006) used 14 items from these two scales, but Edelen et al. (2009) removed five of these items because of lack of variability in response (>90% endorsement of a single response category, p. 1260). Because of the similar results in Orlando et al. (2006) and Edelen et al. (2009), we chose to focus on Edelen et al. (2009) here. In Edelen et al. (2009), two scales are presented: acceptance of female-on-male violence (GB; 5 items, α=.71; e.g., "Boys sometimes deserve to be hit by the girls they date") and acceptance of male-on-female violence (BG; 4 items, α=.55; e.g., "It is OK for a boy to hit a girl if she hit him first"). The factor structure of these scales was explored in Orlando et al. (2006). Higher scores on both scales indicate more acceptance of violence.

Both Edelen et al. (2009) and Orlando et al. (2006) used item response theory (IRT) to explore these scales, due to concerns about non-equivalence (i.e., that gender invariance does not exist for attitude scales); IRT can detect differential item functioning (DIF), which indicates different measurement properties of items by sub-group. As discussed by Edelen et al. (2009), "an item is said to exhibit DIF if two respondents from distinct subgroups who have equal levels of the psychological trait being measured do not have the same probability of endorsing each response category of that item" (p. 1245). Using an item characteristic curve (ICC), location (*b*) and slope (*a*) parameters for each item are determined. The *b* parameter is "the point along the ICC at which the probability of a positive response for a dichotomous item is 50%" and the *a* parameter "represents the slope of the ICC at the value of the location parameter and indicates the extent to which the item is related to the underlying construct" (p. 1247); items with steeper slopes are more discriminating, while a larger location parameter indicates that the participant must have more of a particular construct before endorsing the item. DIF indicates whether there is a difference between the *a* or *b* parameters by sub-group (for a more thorough review, including assessment of model fit, see Orlando et al., 2006). Finally, IRT is also able to handle

highly skewed response distributions, which often occur when measuring attitudes (Orlando et al., 2006).

For the GB scale, two items ("It is OK for a girl to hit a boy if he hit her first" and "Suppose a boy hits a girl, do you think it is wrong for her to hit him back?") had significant *b* DIF (i.e., differing *b* parameters), while for the BG scale, one item had significant *b* DIF ("Suppose a girl hits a boy, do you think it is wrong for him to hit her back?"). The ICCs for these items indicated that given the same overall level of acceptance of GB violence, boys were more accepting of these two GB items than girls, while given the same overall level of acceptance of BG violence, girls were more accepting of this BG item than boys. In both cases, individuals were more accepting when the depicted victim was sex-concordant.

Without considering the item response function, raw response proportions in each category indicated some discriminatory power, although for two of the GB items and three of the BG items, more than half the respondents were in the strongly disagree/really wrong category (see Table 4.1 for scoring). Also, for one of the BG items ("Suppose a girl says something bad to a boy, do you think it is wrong for him to hit her?"), 95.7% of boys and 96.6% of girls said that they thought this was really wrong, indicating poor discriminatory power of this item. Under the IRT model and accounting for DIF, however, estimates for individual item parameters on the GB and BG scales were fairly discriminating, with eight of nine slope estimates exceeding one (which corresponds to a factor loading of ~0.5; Edelen et al., 2009). The item with an *a* parameter of less than one was on the BG scale ("Suppose a girl says something bad to a boy, do you think it is wrong for him to scream at her?"). There was also a wide range of *b* parameter estimates for these items, indicating that items covered a considerable portion of the underlying latent construct continuum (p. 1251); these parameter values also indicated more latent construct coverage for the GB than BG scale.

The presence of DIF can impact scoring; investigating this, Edelen et al. (2009) found that when accounting for DIF, expected total GB scores were higher for boys than girls at all levels of latent girl-on-boy violence acceptance, and expected total BG scores were higher for girls than boys at higher latent levels of boy-on-girl violence acceptance. To assess if scoring that incorporated DIF resulted in different findings than näive, non-DIF scoring, Edelen et al. (2009) calculated observed mean scores, as well as IRT scores that did and did not account for DIF, and compared male versus female means using t-tests. For the GB scale, scores incorporating DIF produced the same pattern of results as non-DIF scores (with boys having significantly higher mean scores for all scoring methods), although the difference between male and female means was more pronounced in non-DIF adjusted scores, which Edelen et al. (2009) concluded was because "when [DIF] [was] ignored in the [observed] and [IRT non-DIF] measures, the DIF [was] conflated with true differences in the genders on the underlying construct and these unadjusted measures exaggerated gender differences" (p. 1256). For the BG scale, the opposite was found, such that only the DIF-scored scale showed a significant difference between the female and male means (with girls having lower mean scores), which Edelen et al. (2009) concluded was because the unadjusted scores "exaggerat[ed] the overall acceptance of boy-on-girl violence by girls and mask[ed] the true gender differences" (p. 1256).

Thus, these two papers demonstrate the importance of considering gender invariance when assessing dating violence attitudes, as the failure to do so can impact empirical findings, and present IRT as one way of assessing this invariance. Since other work presented in Part 2 of this review also found significant differences between males and females (e.g., that females are typically less accepting of violence overall than males), using IRT to detect DIF on other attitude scales seems warranted. Limitations of the Edelen et al. (2009) BG and GB scales include low reliability of the BG scale and limited validity evidence. Also, the two scales that were used to

supply items (Prescribed Norms and NOBAGS) had different response options, which may have resulted in some response differences (e.g., because the NOBAGS scale used the word *wrong* in one of its response options, so participants may have felt more compelled to choose this option; p. 1258). Finally, these scales were assessed in a sample of 9th grade Latino students, and so future work should investigate their properties in more diverse samples, both in terms of age and ethnicity. Particularly, studies investigating DIF by age would be of interest, as this may also impact attitudes (e.g., Price et al., 1999).

Articulated Thoughts in Simulated Situations (ATSS). Rayburn et al. (2007) assessed reactions to dating violence using a think-aloud cognitive paradigm, the ATSS, designed to better understand sex differences in dating violence reactions, as well as how these reactions might differ by sex of described perpetrator and victim.[19] In their prior work, Rayburn et al. (2007) identified a need for a non-self-report attitude scale, as they found that responses to self-report scales were often highly skewed (i.e., participants did not often endorse violence-supportive attitudes), which they thought may have in part resulted from social desirability. Beyond social desirability, however, they also suspected that even if their adolescent participants did generally hold anti-violence attitudes, these attitudes might change if the context of a particular situation changed (e.g., if the described victim was a friend vs. a stranger). Conducting several focus groups with teens, they found that the importance of context in determining attitudes was forefront, leading to the conclusion that "a more meaningful assessment of teens' attitudes about dating violence…[that] incorporated realistic situations from teens' everyday life" was needed (p. 895). The ATSS is such a paradigm, and has previously been used to assess a number of different issues (Davison, Vogel, & Coffman, 1997), including dating violence perpetration by young adults (Eckhardt & Jamison, 2002). As a non-self-report measure, the

[19] The sample used for this assessment was part of the sample used in Schultz and Jaycox (2008; see Part 1), Edelen et al. (2009) and Orlando et al. (2006).

ATSS also has the advantage of not relying on retrospective reporting of general thoughts and attitudes.

In the ATSS paradigm used by Rayburn et al. (2007), participants listened to four scenarios, which each described a dating violence situation (e.g., hearing people arguing at a party). After listening to the description of the situation, the participant then listened to five, short (10-15s) segments, which varied the participant-perpetrator relationship (stranger vs. friend) and perpetrator sex (male vs. female), allowing Rayburn et al. (2007) to explore reaction differences by participant sex, perpetrator sex and perpetrator familiarity. The seriousness of the situation increased in each segment, ending with the use of physical violence. After each segment, the participant had 30s to discuss their thoughts. Open-ended responses were coded using a codebook designed by Rayburn et al. (2007), where one point was awarded for each mention of a particular reaction during a story segment (e.g., ridicule of the perpetrator, justification for violence). Using a random selection of 10 ATSS transcripts (each with four stories), intra-class correlation coefficients between two blind coders were > .70 for all codes. The scores for each of the 20 codes were used as outcome measures by Rayburn et al. (2007); scores for each code were generally not normally distributed, with values concentrated around 0 and 1 (possible range of 0-20), indicating low discriminatory power of the current coding system (although this may also be because of the small sample size, n=41).

After the ATSS, Rayburn et al. (2007) debriefed all participants and performed a manipulation check, which could be considered forms of validity and reliability, respectively. During the debrief, participants reported that they found scenarios quite realistic, that they were able to imagine the stories, and that they had thoughts that were similar to those they would experience if the situation happened in real life. They also found talking aloud about their thoughts and feelings fairly easy. During the manipulation check, the majority of participants

were able to correctly identify whether the perpetrator in the situation had been described as a friend or stranger (85% correct), and whether the victim was portrayed as a friend (not varied by situation; 87% correct).

Rayburn et al. (2007) also present bootstrapped regression models for each of the 20 outcome scores, which included all two- and three-way interactions between participant sex, perpetrator sex, and participant-perpetrator familiarity, and could be considered a form of construct validity. While Rayburn et al. (2007) found numerous associations, and discuss that findings related to differences by participant and perpetrator sex reflect prior findings in non-Latino samples, the number of associations presented and lack of *a priori* hypotheses makes it difficult to summarize and comment on their results. Further, while most associations were as might be expected given past work (e.g., participants were more likely to say they would help the victim escape when the perpetrator was a stranger instead of a friend), some require further exploration (e.g., girls both justified *and* condemned the violence more when the perpetrator was a friend instead of a stranger).

Thus, while this study presents preliminary evidence for the construct validity of the ATSS, research guided by *a priori* hypotheses is needed. Other limitations of this study include the low response rate (to enroll participants for this study, Rayburn et al. (2007) had teachers send home consent materials with 411 students; of these, 55 returned the consent (13.4% return rate) and 41 participated), the restricted age group (9^{th} grade Latino/as) and the small sample size (n=41). Despite this, this study demonstrates the importance of context, time, action and target in understanding attitudes (and therefore potentially behavior; Rayburn et al., 2007, p. 895), and though the ATSS may not be feasible for large samples, indicates ways in which self-administered attitude assessments might be improved.

Empirical Work

We identified six studies that assessed attitudes in males and females using a psychometrically tested scale; all six of these studies used the ATDV scales (Tables 4.2 and 4.3). Of these six studies, three used Canadian participants, with two reporting on data from the same sample (Sears & Byers, 2010; Sears, Byers & Price, 2007); this sample also appears to be the same as the one used to validate the ATDVS. The other three studies used data from a large, representative sample of Thai and Israeli youth, using Thai, Hebrew and Arabic versions of the ATDVS.

Attitude scale means for males are presented in Table 4.2 and for females in Table 4.3. Because of scoring differences, comparisons across the six studies were difficult to make. Specifically, Josephson and Proulx (2008) only reported total AMDV and AFDV scores, calculated using modified scales, and did not report sub-scale scores (in this paper, higher scores indicated more accepting attitudes). Sears and Byers (2010) and Sears et al. (2007) reported only sub-scale scores, with higher scores indicating greater acceptance of dating violence. Pradubmook-Sherer (2011), Sherer (2010) and Pradubmook-Sherer and Sherer (2011) also reported sub-scale scores, but with different scoring than in the two Sears studies. In their Methods section, Pradubmook-Sherer and Sherer (2011) and Pradubmook-Sherer (2011) reported using the same scale scoring as Sears et al. (2007) and Sears and Byers (2010), where 1=*strongly disagree* and 5=*strongly agree* (such that higher scores correspond to attitudes that are more supportive of dating violence), but also stated that "the lower the score, the stronger the endorsement of attitudes supporting dating violence" (Pradubmook-Sherer & Sherer, 2011, p. 818); the discussion of results in the two Pradubmook-Sherer papers indicated that the latter was the case.[20] Also, for all three Pradubmook-Sherer and Sherer articles, it appeared that the reported scores were a mean of mean sub-scale scores, and not a mean of sum scores (since a

[20] The opposite seemed to be the case in Sherer (2010), such that higher scores in this paper indicated more violence-supportive attitudes, and so results from Pradubmook-Sherer and Sherer (2011) were reported instead of results from Sherer (2010), in order to be consistent with results from Pradubmook-Sherer (2011).

mean of the sum score was out of range). Thus, the six listed studies were not directly comparable, although all indicated that boys had attitudes that were more supportive of violence than girls, except for female scores on the AFDV-Phys and AMDV-Psyc scales in Sears et al. (2007) and Sears & Byers (2010), respectively, which were slightly higher (i.e., more supportive) of dating violence than male scores. Results from Pradubmook-Sherer and Sherer (2011) also indicated within-sex cultural differences, such that male and female Arab and Thai youth had lower (i.e., more supportive) scores than Jewish youth, which Pradubmook-Sherer and Sherer (2011) hypothesize is because of the more traditional cultures of the former. Thus, additional cross-cultural research is of interest; future work should also consider the use of other attitude measures (e.g., AADS, JVCT) that have greater discriminatory power than the ATDVS.

Table 4.2
Male Attitudes

Authors and year	Measure	Number of items	N	% female	Race/ ethnicity	Mean (SD)
Community-based samples						
Josephson & Proulx (2008)[1]	ATDVS – short form	36 (6 per sub-scale)	290	50.7[a]	Western Canadian	AMDV: 11.45 (4.06)[b] AFDV: 13.54 (3.26)[b]
Sears & Byers (2010)[1]	AFDV	37 (12 P, 13 Ps, 12 S)	627[c]	49.4[a]	Primarily White	AFDV – Psyc: 28.44 (7.46) AFDV – Phys: 29.95 (9.18) AFDV – Sex: 26.46 (10.02)
Sears et al. (2007)[1]	AMDV	39 (12 P, 15 Ps, 12 S)	633[c]	48.8[a]	Primarily White	AMDV – Psyc: 32.58 (10.14) AMDV – Phys: 23.34 (9.48) AMDV – Sex: 22.60 (9.90)
Probability/census samples						
Pradubmook-Sherer (2011)[3]	AMDV – Thai	39 (12 P, 15 Ps, 12 S)	1296[d]	54.9[a]	Thai	AMDV – Psyc: 3.25 (0.46) AMDV – Phys: 3.00 (0.48) AMDV – Sex: 3.30 (0.74)
Sherer (2010) and Pradubmook-Sherer & Sherer (2011)[2]	AMDV – Hebrew & Arabic	39 (12 P, 15 Ps, 12 S)	1357[e]	56.4[a]	Israeli (52.8% Arab, 47.2% Jewish)[a]	AMDV – Psyc: 2.90 (0.092), 3.05 (0.060) AMDV – Phys: 2.91 (0.12), 3.40 (0.050) AMDV – Sex: 3.13 (0.088), 3.41 (0.050)
Pradubmook-Sherer & Sherer (2011)[3]	AMDV – Thai	39 (12 P, 15 Ps, 12 S)	616[f]	58.0	Thai	AMDV – Psyc: 2.85 (0.03) AMDV – Phys: 2.80 (0.02) AMDV – Sex: 3.29 (0.04)

Abbreviations: P – Physical; Ps – Psychological; S – Sexual
[1]Location: Canada
[2]Location: Israel
[3]Location: Thailand

[a]Hand calculated
[b]Scores across the three sub-scales were combined and averaged.
[c]Sears et al. (2007) and Sears & Byers (2010) use the same sample, but present different results.
[d]This study randomly sampled adolescents aged 14 to 19 (*M* age ~16) in the Bangkok, Thailand area. Participants were recruited from high schools, vocational

schools, and out-of-school settings.
[e]Pradubmook-Sherer & Sherer (2011) and Sherer (2010) use the same sample. Results are presented from Pradubmook-Sherer and Sherer (2011). Values are for Arab means and Jewish means, respectively.
[f]Same sample as Pradubmook-Sherer (2011), but includes high school students only.

Table 4.3
Female Attitudes

Authors and year	Measure	Number of items	N	% female	Race/ ethnicity	Mean (SD)
Community-based samples						
Josephson & Proulx (2008)[1]	ATDVS – short form	36 (6 per sub-scale)	290	50.7[a]	Western Canadian	AMDV: 9.57 (3.01)[b] AFDV: 11.64 (3.69)[b]
Sears & Byers (2010)[1]	AMDV	39 (12 P, 15 Ps, 12 S)	627[c]	49.4[a]	Primarily White	AMDV – Psyc: 28.68 (7.59) AMDV – Phys: 20.09 (7.73) AMDV – Sex: 17.18 (5.89)
Sears et al. (2007)[1]	AFDV	37 (12 P, 13 Ps, 12 S)	633[c]	48.8[a]	Primarily White	AFDV – Psyc: 27.39 (7.90) AFDV – Phys: 26.40 (9.69) AFDV – Sex: 18.37 (6.35)
Probability/census samples						
Pradubmook-Sherer (2011)[3]	AMDV – Thai	39 (12 P, 15 Ps, 12 S)	1296[d]	54.9[a]	Thai	AMDV – Psyc: 3.70 (0.45) AMDV – Phys: 3.39 (0.49) AMDV – Sex: 3.84 (0.62)
Sherer (2010) and Pradubmook-Sherer & Sherer (2011)[2]	AMDV – Hebrew & Arabic	39 (12 P, 15 Ps, 12 S)	1357[e]	56.4[a]	Israeli (52.8% Arab, 47.2% Jewish)[a]	AMDV – Psyc: 2.30 (0.054), 3.56 (0.054) AMDV – Phys: 3.15 (0.096), 3.88 (0.060) AMDV – Sex: 3.85 (0.059), 4.15 (0.045)
Pradubmook-Sherer & Sherer (2011)[3]	AMDV – Thai	39 (12 P, 15 Ps, 12 S)	616[f]	58.0	Thai	AMDV – Psyc: 3.29 (0.02) AMDV – Phys: 2.93 (0.03) AMDV – Sex: 3.74 (0.04)

Abbreviations: P – Physical; Ps – Psychological; S – Sexual
[1]Location: Canada
[2]Location: Israel
[3]Location: Thailand

[a]Hand calculated
[b]Scores across the three sub-scales were combined and averaged.
[c]Sears et al. (2007) and Sears & Byers (2010) use the same sample, but present different results.
[d]This study randomly sampled adolescents aged 14 to 19 (*M* age ~16) in the Bangkok, Thailand area. Participants were recruited from high schools, vocational

schools, and out-of-school settings.
[e]Pradubmook-Sherer & Sherer (2011) and Sherer (2010) use the same sample. Results are presented from Pradubmook-Sherer and Sherer (2011). Values are for Arab means and Jewish means, respectively.
[f]Same sample as Pradubmook-Sherer (2011), but includes high school students only.

Summary: Part 2

In our searches, seven dating violence attitude measures with evidence of reliability and validity were located: the ATDVS (AMDV and AFDV), the AADS, the JVCT, the AIV, the AARDR, the Prescribed Norms/NOBAGS scale (BG and GB) and the ATSS. Six of these scales assess attitudes toward physical aggression, five assess attitudes toward psychological aggression and two assess attitudes toward sexual aggression. One scale, the AARDR, also contains items that tap attitudes about healthy dating relationships. Overall, these scales were more varied in content and approach than the behavior measures described in Part 1. Located empirical articles all used the ATDVS, but findings across these articles were not directly comparable because of modifications to scale items and scoring.

General Discussion (Parts 1 and 2)

In the two parts of this review, we identified 13 teen dating violence measures with evidence of reliability and validity, six that assess behaviors and seven that assess attitudes. The majority of these measures were tested in non-representative samples and did not present direct information on discriminatory power; the former finding also limits the development of norms. Of those providing reliability information (all but one article), 11 of 12 assessed internal consistency reliability, with a smaller number also considering test-retest reliability. Of those assessing validity (all but two articles), the most commonly assessed form was construct (65.6% of all validity assessments), with the majority of these articles assessing convergent validity (42.9%) or factor structure (28.6%), and fewer assessing discriminant validity, measurement invariance or general construct validity. A smaller number of articles included assessments of concurrent validity (9.4% of all validity assessments), criterion validity (9.4%), content validity (6.3%) or predictive validity (3.1%). The mean number of specific validity investigations per article was one (range, 0-3).

General Observations: Comparison to Evidence-Based Programs

During the process of locating these 13 measures, it became clear that a large number of existing empirical articles used measures that were not tested, had no evidence of reliability beyond internal consistency and had no evidence of validity—the quality of measure development for these untested measures varied widely. This is not unlike the large number of public health programs that are currently offered, but which have never received any formal testing to determine their effectiveness. Rather, implementers of these programs rely on the belief that the program works, perhaps because they feel good when they offer the program, because they designed the program, or because participants report liking the program (a face validity of sorts). However, to ensure that programs are likely to produce the target outcomes, the selection and implementation of evidence-based programs, or programs with strong evidence of effectiveness, is now encouraged or required by many organizations. To facilitate this selection, a number of tools exist to help practitioners select the best program for their needs (e.g., the National Registry of Evidence Based Programs, Blueprints for Violence Prevention). The present compendium is a first step in establishing a tool that helps researchers and practitioners choose teen dating violence measures that meet their needs *and* have evidence of effectiveness, which is as important to strong empirical research as use of evidence-based programs is to strong public health impacts.

The empirical portion of this review identified a limited number of articles that used selected measures, and those that were identified sometimes used a modified version of an existing scale. Returning to the evidence-based program analogy, this seems not unlike the fidelity vs. adaptation debate that surrounds these programs (Blakely et al., 1987; Cohen et al., 2008), as researchers likely modified measures because they felt the existing, tested measure was not a "fit" for their study (e.g., because it was too long, because it was not culturally relevant);

however, just as in program adaptation, careful consideration should be given before adapting an evidence-based measure, as adaptations limit the comparability of the new study with existing research, and may also threaten reliability, validity, sensitivity and specificity.

Similar to many evidence-based programs (which were only tested in one population at one time), the amount of evidence supporting the reliability and validity of these 13 measures is limited. Of the 13 reviewed measures, only four were assessed in more than one study, and only the CADRI was assessed in more than two studies. Thus, for behavioral measures, the CADRI has the most robust support of its reliability and validity, although these numerous studies also identified some problems with this measure (e.g., with the sexual abuse sub-scale). Of the attitude measures, the ATDVS, AADS and JVCT are the most studied, with the AADS specifically designed to increase discriminatory power. However, innovative research by Edelen et al. (2009) and Orlando et al. (2006) demonstrates that assessment of gender invariance is needed for all three of these attitude measures. Thus, no one measure was identified as the "gold standard" by this review. Instead, researchers are encouraged to use provided information about the strengths and limitations of each measure to select the scale that best meets their needs.

Promising Measures

While the goal of this review was to identify "evidence-based" measures, defined as measures that were the focus of systematic psychometric testing, several other promising measures were also identified (i.e., measures that were the subject of preliminary investigations of reliability and validity). These five measures are now reviewed briefly (three were located in computer-based searches, one was located in computer-based and review article searches, and the other was given to the first author by another researcher).

Safe Dates. Although Foshee's Safe Dates scales (e.g., Foshee et al., 1998) are featured in the CDC compendia discussed in the Part 1 Introduction, they have not been the focus of

specific psychometric testing, although their use over time has amassed evidence that supports reliability and validity, such as correlation patterns with variables of interest (e.g., Foshee, Linder, MacDougall, & Bangdiwala, 2001; Foshee et al., 2005; Foshee et al., 2007; V. A. Foshee, personal communication, October 29, 2012). In terms of the development of these scales, one article from this research group elaborates that at the time Safe Dates data collection started "previously validated scales were not available for measuring several of the proposed theoretical constructs such as adolescent dating violence…thus, we created scales to measure those constructs specifically for this study through several processes, including interviewing adolescents about their interpretation of item meaning, administering draft scales to adolescents in school classrooms, conducting psychometric analyses, conducting Q-sort procedures, and readministering revised scales to a separate sample of 300 adolescents" (Foshee et al., 2001, p. 132). Given the popularity of these scales, work that summarized the various steps taken to develop the Safe Dates measures, as well as presented a summary of evidence supporting reliability and validity, would benefit the field.

Identification, acceptance and exposure to violence. Bobowick (2000) developed scales assessing identification and acceptance of, and direct and indirect exposure to, dating violence, using a small sample (n=45) of high school freshman and seniors from Oregon. Scales were developed by reviewing the literature (the direct exposure scale was adapted from the CTS), with expert review conducted by two psychologists. Pilot testing was performed with nine high school students. Internal consistency reliabilities for final scales were low to adequate. Specific hypotheses about these scales included that participants would identify a relationship as abusive more frequently when the violence was described as physical and sexual, as opposed to verbal; that there would be a positive association between dating violence and exposure to violence; and that there would be a positive association between acceptance of dating violence

and risk-taking behavior (pp. 26-27). Most hypotheses were not supported, but Bobowick (2000) did find that participants more readily identified and found more acceptable male-to-female physical abuse than female-to-male physical abuse, which coincides with previous literature. Additional work with an adequately powered sample is needed, with hypotheses and analyses that specifically focus on establishing validity, but this measure is promising because of its comprehensive assessment of attitudes and behaviors, as well as the presence of open- and closed-ended response options; open-ended responses allow for more nuanced investigations of perceptions of risk and consequences of different types of dating violence (e.g., if male-to-female violence is viewed as more dangerous than female-to-male violence).

Theory of Planned Behavior scales. For an empirical study of dating violence in a South African sample, Flisher, Myer, Mèrais, Lombard and Reddy (2007) developed several items based on the Theory of Planned Behavior; these items assess attitudes toward partner violence (seven items), subjective norms about partner violence (seven items), outcome expectancy about partner violence (two items), and self-efficacy regarding partner violence prevention (one item; these scales/items are collectively referred to by Flisher et al. (2007) as the predictor variables). Items were administered to a random sample of grade 8 and 11 students in Cape Town, South Africa (n=596). Two-week test-retest reliability (using data from a sub-sample, n=114) demonstrated somewhat poor stability (r range .34-.59), and internal consistency reliabilities for multi-item scales, as indicated by Cronbach's alpha, were low (.58-.65). Controlling for all predictor variables, as well as age and sex, Flisher et al. (2007) found that attitudes and general norms about partner violence significantly predicted intentions to perpetrate physical violence, while attitudes and outcome expectancies about partner violence significantly predicted actual use of physical violence, partially supporting the hypothesis that all predictor variables would be associated with intentions to use, and actual use of, violence. Thus, while this

study did not focus on establishing the validity of these predictor variables, work in this article provides preliminary evidence for the construct validity of some scales. Other important features of these items include their theoretical grounding, and their availability in different languages (English, Afrikaans and Xhosa).

Adolescent Dating Measure. Pyle (1990) details the development of the Adolescent Dating Measure (ADM), intended to assess females' negative dating experiences. Items were developed through interviews with 12 adolescent and five adult women and from the investigator's prior work experience; items were pilot-tested on seven adolescent and six adult women. The scale contains 92 items, which primarily assess psychologically aggressive actions and are comprehensive in scope. Split-half reliability for these items in Pyle's (1990) adolescent testing sample was .90. On the ADM, respondents are asked to indicate not only if the behavior has happened to them, but what their response to the behavior was (e.g., caused me to terminate the relationship). In a sample of 106 Californian females aged 15-17, Pyle (1990) used ADM responses to develop a weighted scoring system and to determine normative levels of dating violence, in order to facilitate comparison of retrospective reports of teen dating violence by a sample of battered and non-battered adult women. The weighted scoring system resulted in an approximately normal distribution of weighted frequency counts, indicating that a weighting system accounting for behavioral frequency and response may be one way to increase discriminatory power of behavioral items; however, this concept should be applied to a measure with evidence of reliability and validity and using a representative sample.

Psychological Maltreatment of Women Inventory – Teen. Mac Neil (2010) reports on a 30-item adolescent version of the Psychological Maltreatment of Women Inventory (PMWI-Teen). An adolescent version of the PMWI was first introduced by Price (2002); Mac Neil (2010) added contextual prompts to this version, where respondents indicate the most stressful

experience from the 30 items, and then provide detailed information on this experience (e.g., how long ago it happened, who was present when it happened). Internal consistency for the PMWI-Teen was high (α=.93) in a sample of 168 grade 10-12 Canadian students. In what could be considered preliminary evidence of construct validity, Mac Neil (2010) hypothesized that more than 50% of participants would report experiences with psychological aggression, that girls would be more likely to report these behaviors than boys, and that more frequent experiences of psychological aggression would be associated with more symptoms of depression and anxiety and more substance use and misconduct. Mac Neil (2010) also hypothesized that relationships with these outcome variables would be stronger for adolescents who perceived the psychological aggression as stressful, and weaker for adolescents who perceived the psychological aggression as less stressful (e.g., more controllable), but he did not have specific hypotheses about which actions or contexts would be seen as most stressful (pp. 41-44). In this sample, 99% of participants reported at least one experience with psychological aggression, and other hypotheses were partially supported. Thus, although this study was not intended to validate the PMWI-Teen, it suggests that it is a promising measure for future study; other topics for future study include adolescent feedback on item content, exploration and confirmation of factor structure, and analyses that use the contextual information to improve discriminatory power.

Limitations of Current Knowledge and Future Directions

In sum, there exist a number of measures of teen dating violence with some evidence of reliability and validity, but these measures are not commonly used in empirical study to assess prevalence or attitudes. Evidence of the quality of these tested measures varies, with none designated as a gold standard for teen dating violence measurement. Work on the measurement of teen dating violence is needed, and funding solicitations that focus on this measurement are a positive step, with future funding solicitations of this nature encouraged. Teen dating violence

measurement is also a topic with the potential for innovative research that will contribute not only to the advancement of this field, but also to broader understandings of adolescent attitude and behavior measurement.

Considering possible directions for this future research, both Hilton, Harris and Rice (1998) and Jouriles et al. (2005) found that extending the recall time frame had little effect on frequency reports of interpersonal violence among adolescents (e.g., same frequency reported whether the recall period was two weeks or two months), and assessing general aggression in an adolescent sample, Hilton et al. (2003) found that "memory for specific behaviors was quite imperfect after only 10 minutes" (p. 234; participants were asked to recall behaviors they had heard in an aggression scenario). Although participants were not recalling events that had necessarily happened to them, Hilton et al. (2003) argue that "the retention interval…was so short and performance so far from perfect that the results, in our opinion, raise concerns about the accuracy of participants' reports of actual experiences" (p. 234). Since most dating violence research is retrospective in nature, issues of recall ability and reporting accuracy are critical to empirical study. Work in cognition and memory has demonstrated that memory of past events is fallible (Offer, Kaiz, Howard, & Bennett, 2000), and that while memories of positive and negative events may be more vivid than memories of neutral events (Hamann, 2001; Ensinger & Corkin, 2003), positive memories may be remembered in more detail than negative memories (D'Argembeau, Comblain, & Van Der Linden, 2003), such that negative events are less accessible to recall (Taylor, 1991). Both positive and negative memories are also subject to distortion (though negative memories may be less distortion prone; Kensinger & Schacter, 2006), and sex differences in recall have also been documented (Seidlitz & Diener, 1998; Hilton, Harris, & Rice, 2003). Further, experiences that are more subtle in nature and require more judgment and interpretation, as is potentially the case with many teen dating violence items, are subject to

especially poor recall (Hardt & Rutter, 2004). While a full review of memory research is beyond the scope of this paper, this prior work demonstrates memory recall issues that have important implications for teen dating violence measurement, and are fruitful topics for future research.

More broadly, future research should also focus on more rigorously defining the construct of teen dating violence, including a consideration of how dating itself is characterized; as discussed by Teten, Ball, Valle, Noonan and Rosenbluth (2009), the inability of current measures to fully capture the construct may be one source of mixed empirical findings. Follingstad and Rogers (2013) also question whether physical, sexual and psychological aggression can be validly measured using the same method, which in part depends on what is included in the construct of each behavior. Future measurement work should also determine how to incorporate context (e.g., perceived intent, recipient's reaction) to obtain more nuanced and realistic pictures of dating violence; several measures discussed in this review have begun work in this direction (e.g., Jouriles et al., 2005; Leisen, 2000; Mac Neil, 2010). In sum, more comprehensive and thoughtful approaches to defining physical, sexual and psychological aggression, as well as attitudes toward aggression, are important topics for research and discussion.

Limitations of the Review

While this review indicates that teen dating violence measurement is an under-developed area, with many opportunities for new research, limitations should be noted. First, selecting measures to include in this type of review is a subjective process (Rathus & Feindler, 2004). While we used specific inclusion criteria to be as objective as possible, this did not eliminate all subjectivity (for example, choosing between evidence-based and promising measures). To address this, we have presented detailed information on all included measures, as well as on measures we deemed promising, in order to be as transparent as possible, and allow room for

debate and discussion. Indeed, we hope this article is a first step to more vigorous debate and interest in teen dating violence measurement. Second, as noted in Part 1 of this review, our review of the empirical literature was not systematic, but rather aimed to locate a representative sample of empirical articles that used the 13 included measures over the past ten years. While this targeted search approach increased feasibility, and also allowed the search to be focused only on included measures, we acknowledge that this method may have missed some empirical articles. However, we feel the results of our search support the conclusion that a substantial number of dating violence prevalence estimates were not obtained with evidence-based measures. Finally, we opted not to provide a summary table that scored various measures on their quality, as we felt this would detract from the purpose of the review, and is also a questionable practice for validity assessments (since a measure is not "valid" in the same way a measure can be reliable). Our approach instead has the advantage of allowing readers to assess the available evidence and draw conclusions about each measure, but the disadvantage of not providing quick summaries of reviewed measures. However, we felt the latter would not be particularly informative and could be misleading, while the current approach provides the reader with information needed to assess the state of teen dating violence measurement, and more specifically, chose a measure for their research, both of which were goals of the review.

Summary

What is measured determines what is found (Cook, 2002); while this may seem obvious, its implications can be discounted when need for information on a pressing public health problem is great, as was the case at the beginning of this field. Now thirty years into empirical work on teen dating violence, the findings of this review suggest the field needs to take a step back, assess what we aim to measure, and determine whether our current measures meet that goal. The 13 measures presented in these two papers will inform this discussion, but all have

room for improvement and require further testing. The development of new measures that better assess the complexity of interpersonal interactions, and consider how to minimize the biases of retrospective reporting and/or find new and innovative ways to assess dating violence, is also an important goal. With work that focuses on improving measurement, empirical research produced in the next thirty years will more precisely inform and shape public health priorities.

References (Parts 1 and 2)

Ackard, D. M., Eisenberg, M. E., & Neumark-Sztainer, D. (2007). Long-term impact of adolescent dating violence on the behavioral and psychological health of male and female youth. *Journal of Pediatrics, 151*, 476-481.

Antônio, T., & Hokoda, A. (2009). Gender variations in dating violence and positive conflict resolution among Mexican adolescents. *Violence and Victims, 24*, 533-545.

Archer, J., Fernández-Fuertes, A. A., & Thanzami, V. L. (2010). Does cost-benefit analysis or self-control predict involvement in two forms of aggression? *Aggressive Behavior, 36*, 292-304.

Bandyopadhyay, A., Deokar, A., & Omar, H. A. (2010). Adolescent dating violence: A comprehensive review. *International Journal of Child and Adolescent Health, 3*, 305-320.

Black, B. M., Tolman, R. M., Callahan, M., Saunders, D. G., & Weisz, A. N. (2008). When will adolescents tell someone about dating violence victimization? *Violence Against Women, 14*, 741-758.

Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253-268.

Boafo, I. M. (2011). Dating violence and condom use self-efficacy: A study of school-going adolescents in Cape Town. *IFE PsychologIA: An International Journal, 19*, 296-320.

Bobowick, M. *Adolescents' identification, acceptance and response to dating violence*. (Unpublished doctoral dissertation). The California School of Professional Psychology, Alameda, CA.

Brooks-Russell, A., Foshee, V. A., & Ennett, S. T. (2013). Predictors of latent trajectory classes of physical dating violence victimization. *Journal of Youth and Adolescence, 42*, 566-580.

Brown, A., Cosgrave, E., Killackey, E., Purcell, E., Buckby, J., & Yung, A. R. (2009). The longitudinal association of adolescent dating violence with psychiatric disorders and functioning. *Journal of Interpersonal Violence, 24*, 1964-1979.

Bulmer, M. G. (1979). *Principles of statistics*. Cambridge, MA: MIT Press.

Burcky, W., Reuterman, N., & Kopsky, S. (1988). Dating violence among high school students. *The School Counselor, 35*, 353-358.

Buzy, W. M., McDonald, R., Jouriles, E. N., Swank, P. R., Rosenfield, D., Shimek, J. S., & Corbitt-Shindler, D. (2004). Adolescent girls' alcohol use as a risk factor for relationship violence. *Journal of Research on Adolescence, 14*, 449-470.

Callahan, M. R., Tolman, R. M., & Saunders, D. G. (2003). Adolescent dating violence victimization and psychological well-being. *Journal of Adolescent Research, 18*, 664-681.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Capaldi, D. M., Knoble, N. B., Shortt, J. W., & Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner Abuse, 3*, 231-280.

Cascardi, M., Avery-Leaf, S., O'Leary, K. D., & Slep, A. M. S. (1999). Factor structure and convergent validity of the Conflict Tactics Scale in high school students. *Psychological Assessment, 11*, 546-555.

Centers for Disease Control and Prevention [CDC]. (2012). Youth risk behavior surveillance – United States, 2011. *Morbidity and Mortality Weekly Report, 61*, 1-162.

Cohen, D. J., Crabtreee, B. F., Etz, R. S., Balasubramanian, B. A., Donahue, K. E., Leviton, L. C., … Green, L. W. (2008). Fidelity versus flexibility: Translating evidence-based research into practice. *American Journal of Preventive Medicine, 35*, S381-S389.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence

Erlbaum Associates.

Collin-Vézina, D., Hébert, M., Manseau, H., Blais, M., & Fernet, M. (2006). Self-concept and dating violence in 220 adolescent girls in the child protective system. *Child & Youth Care Forum, 35*, 319-326.

Connolly, J., Nocentini, A., Menesini, E., Pepler, D., Craig, W., & Williams, T. S. (2010). Adolescent dating aggression in Canada and Italy: A cross-national comparison. *International Journal of Behavioral Development, 34*, 98-105.

Cook, S. L. (2002). Self-reports of sexual, physical, and nonphysical abuse perpetration: A comparison of three measures. *Violence Against Women, 8*, 541-565.

Cramer, D. (1997). *Basic statistics for social research: Step-by-step calculations and computer techniques using Minitab*. New York, NY: Routledge.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent from psychopathology. *Journal of Consulting and Clinical Psychology, 24*, 349-354.

Cyr, M., McDuff, P., & Wright, J. (2006). Prevalence and predictors of dating violence among adolescent female victims of child sexual abuse. *Journal of Interpersonal Violence, 21*, 1000-1017.

Dahlberg, L. L., Toal, S. B., Swahn, M., & Behrens, C. B. (2005). *Measuring violence-related attitudes, behaviors, and influences among youths: A compendium of assessment tools* (2nd Ed.). Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

D'Argembeau, A., Comblain, C., & Van Der Linden, M. (2003). Phenomenal characteristics of autobiographical memories for positive, negative, and neutral events. *Applied Cognitive*

Psychology, 17, 281-294.

Davidson, M. M. (2005). *Adolescent attitudes regarding dating relationships: The construction of sex-specific scales.* (Unpublished doctoral dissertation). University of Missouri-Columbia, Columbia, MO.

Davison, G. C., Vogel, R. S., & Coffman, S. G. (1997). Think-aloud approaches to cognitive assessment and the articulated thoughts simulated situations paradigm. *Journal of Consulting & Clinical Psychology, 65*, 950-958.

Eckhardt, C., & Jamison, T. R. (2002). Articulated thoughts of male dating violence perpetrators during anger arousal. *Cognitive Therapy & Research, 26*, 289-308.

Edelen, M. O., McCaffrey, D. F., Marshall, G. N., & Jaycox, L. H. (2009). Measurement of teen dating violence attitudes: An item response theory evaluation of differential item functioning according to gender. *Journal of Interpersonal Violence*, *24*, 1243-1263.

Emelianchik-Key, K. M. (2010). *The initial development and validation of the Teen Screen for Dating Violence*. (Unpublished doctoral dissertation). Old Dominion University, Norfolk, VA.

Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure, and results. *International Journal of Testing, 10*, 295-317.

Exner-Cortens, D., Eckenrode, J., & Rothman, E. (2013). Longitudinal associations between teen dating violence victimization and adverse health outcomes. *Pediatrics*, *131*, 71-78.

Fernández-Fuertes, A. A., & Fuertes, A. (2010). Physical and psychological aggression in dating relationships of Spanish adolescents: Motives and consequences. *Child Abuse & Neglect, 34*, 183-191.

Fernández-Fuertes, A. A., Fuertes, A., & Pulido, R. F. (2006). Evaluación de la violencia en las

relaciones de pareja de los adolescents. Validación del *Conflict in Adolescent Dating Relationships Inventory (CADRI)* – version española. *International Journal of Clinical Health and Psychology, 6,* 339-358.

Fernández-González, L., Wekerle, C., & Goldstein, A. L. (2012). Measuring adolescent dating violence: Development of 'Conflict in Adolescent Dating Relationships Inventory' short form. *Advances in Mental Health, 11*, 35-54.

Flisher, A. J., Myer, L., Mèrais, A., Lombard, C., & Reddy, P. (2007). Prevalence and correlates of partner violence among South African adolescents. *Journal of Child Psychology and Psychiatry, 48*, 619-627.

Follingstad, D. R., & Rogers, M. J. (2013). Validity concerns in the measurement of women's and men's report of intimate partner violence. *Sex Roles, 69*, 149-167.

Foshee, V. A., Bauman, K., Arriaga, X., Helms, R., Koch, G., & Linder, G. F. (1998). An evaluation of Safe Dates, an adolescent dating violence prevention program. *American Journal of Public Health, 88*, 45-50.

Foshee, V. A., Bauman, K. E., Linder, F., Rice, J., & Wilcher, R. (2007). Typologies of adolescent dating violence: Identifying typologies of adolescent dating violence perpetration. *Journal of Interpersonal Violence, 22*, 498-519.

Foshee, V. A., Bauman, K. E., Ennett, S. T., Suchindran, C., Benefield, T., & Linder, G. F. (2005). Assessing the effects of the dating violence prevention program "Safe Dates" using random coefficient regression modeling. *Prevention Science, 6*, 245-258.

Foshee, V. A., Benefield, T., Suchindran, C., Ennett, S. T., Bauman, K. E., Karriker-Jaffe, K. J., … Mathias, J. (2009). The development of four types of adolescent dating abuse and selected demographic correlates. *Journal of Research on Adolescence, 19*, 380-400.

Foshee, V. A., Linder, G. F., Bauman, K. E., Langwick, S. A., Arriaga, X. B., Health, J. L., . . .

Bangdiwala, S. (1996). The Safe Dates project: Theoretical basis, evaluation design, and selected baseline findings. *American Journal of Preventive Medicine, 12*, 39-47.

Foshee, V. A., Linder, F., MacDougall, J. E., & Bangdiwala, S. (2001). Gender differences in the longitudinal predictors of adolescent dating violence. *Preventive Medicine, 32*, 128-141.

Foshee, V. A., & Matthew, R. A. (2007). Adolescent dating abuse perpetration: A review of findings, methodological limitations, and suggestions for future research. In D. J. Flannery, A. T. Vazsonyi, & I. D. Waldman (Eds.), *The Cambridge handbook of violent behavior and aggression* (pp. 431-449). New York, NY: Cambridge University Press.

Foshee, V. A., & Reyes, H. L. M. (2011). Dating abuse: Prevalence, consequences, and predictors. In R. J. R. Levesque (Ed.), *Encyclopedia of Adolescence* (pp. 602-615). New York, NY: Springer Science + Business Media.

Gagné, M. H., Lavoie, F., & Hébert, M. (2005). Victimization during childhood and revictimization in dating relationships in adolescent girls. *Child Abuse & Neglect, 29*, 1155-1172.

Galambos, N. L., Peterson, A. C., Richards, M., & Gitelson, I. B. (1985). The Attitudes Towards Women Scale for Adolescents (AWSA): A study of reliability and validity. *Sex Roles, 13*, 343-356.

Giordano, P. C., Soto, D. A., Manning, W. D., & Longmore, M. A. (2010). The characteristics of romantic relationships associated with teen dating violence. *Social Science Research, 39*, 863-874.

Gray, H. M., & Foshee, V. (1997). Adolescent dating violence: Differences between one-sided and mutually violent profiles. *Journal of Interpersonal Violence, 12*, 126-141.

Halpern, C. T., Oslak, S. G., Young, M. L., Martin, S. L., & Kupper, L. L. (2001). Partner violence among adolescents in opposite-sex romantic relationships: Findings from the

National Longitudinal Study of Adolescent Health. *American Journal of Public Health, 91*, 1679-1685.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *TRENDS in Cognitive Sciences, 5*, 394-400.

Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry, 45*, 260-273.

Hays, D. G., & Emelianchik, K. (2009). A content analysis of intimate partner violence assessments. *Measurement and Evaluation in Counseling and Development, 42,* 139-153.

Henton, J., Cate, R., Koval, J., Lloyd, S., & Christopher, S. (1983). Romance and violence in dating relationships. *Journal of Family Issues, 4*, 467-482.

Hickman, L. J., Jaycox, L. H., & Aronoff, J. (2004). Dating violence among adolescents: Prevalence, gender distribution, and prevention program effectiveness. *Trauma, Violence, & Abuse, 5*, 123-142.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (1998). On the validity of self-reported rates of interpersonal violence. *Journal of Interpersonal Violence, 13*, 58-72.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (2003). Correspondence between self-report measures of interpersonal aggression. *Journal of Interpersonal Violence, 18*, 223-239.

Hokoda, A., Ramos-Lira, L., Celaya, P., Vilhauer, K., Angeles, M., Ruíz, S., … Mora, M. D. (2006). Reliability of translated measures assessing dating violence among Mexican adolescents. *Violence and Victims, 21*, 117-127.

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, *72*, 408-419.

Jackson, S. M., Cram, F., & Seymour, F. W. (2000). Violence and sexual coercion in high school

students' dating relationships. *Journal of Family Violence, 15*, 23-36.

Johnson, B.T., & Eagly, A.H. (2000). Quantitative synthesis of social psychological research. In H.T. Reis, & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 496-528). London: Cambridge University Press.

Josephson, W. L., & Proulx, J. B. (2008). Violence in young adolescents' relationships: A path model. *Journal of Interpersonal Violence, 23*, 189-208.

Jouriles, E. N., Garrido, E., Rosenfield, D., & McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress. *Child Abuse & Neglect, 2009,* 451-460.

Jouriles, E. N., McDonald, R., Garrido, E., Rosenfield, D., & Brown, A. S. (2005). Assessing aggression in adolescent romantic relationships: Can we do it better? *Psychological Assessment, 17*, 469-475.

Kasian, M., & Painter, S. L. (1992). Frequency and severity of psychological abuse in a dating population. *Journal of Interpersonal Violence, 7*, 350-364.

Kernsmith, P. D., & Tolman, R. M. (2011). Attitudinal correlates of girls' use of violence in teen dating relationships. *Violence Against Women, 17*, 500-516.

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*, 1169-1180.

Kensinger, E. A., & Schacter, D. L. (2006). When the Red Sox shocked the Yankees: Comparing negative and positive memories. *Psychonomic Bulletin & Review, 13*, 757-763.

Kline, P. (2000). *A psychometrics primer*. London, UK: Free Association Books.

Knoble, N. B., Capaldi, D. M., Shortt, J. W., & Kim, H. K. (2012). *Risk factors for adolescent dating violence: A systematic review.* Poster presented at the 14th Society for Research on

Adolescence Biennial Meeting, Vancouver, BC.

Lavoie, F., Hébert, M., Tremblay, R., Vitaro, F., Vézina, L., & McDuff, P. (2002). History of family dysfunction and perpetration of dating violence by adolescent boys: A longitudinal study. *Journal of Adolescent Health, 30*, 375-383.

Lavoie, F., & Vézina, L. (2001). Violence faite aux filles dans le contexte des fréquentations à l'adolescence: Élaboration d'un instrument (VIFFA). *Canadian Journal of Community Mental Health/Revue Canadienne de Santé Mentale Communautaire, 20*, 153-171.

Leisen, M. B. (1999). *Development and validation of the Adolescent Partner Aggression Scale (APAS).* (Unpublished doctoral dissertation). University of Hawai'i at Mānoa, Honolulu, HI.

Lichter, E. L., & McCloskey, L. A. (2004). The effects of childhood exposure to marital violence on adolescent gender-role beliefs and dating violence. *Psychology of Women Quarterly, 28*, 344-357.

Lonsway, K. A., & Fitzgerald, L. F. (1995). Attitudinal antecedents of rape myth acceptance: A theoretical and empirical re-examination. *Journal of Personality and Social Psychology, 68*, 701-711.

Lozano, M. P. G. (2009). *Violencia en las relaciones de noviazgo entre jóvenes y adolescentes de la Comunidad de Madrid.* (Unpublished doctoral dissertation). Universidad Complutense de Madrid, Madrid, Spain.

Mac Neil, B. A. (2010). *Adolescents' experiences of psychologically aggressive behavior in their dating relationships and negative emotional and behavioral adjustment.* (Unpublished doctoral dissertation). University of New Brunswick, Saint John, New Brunswick.

Maher, B. A., & Gottesman, I. I. (2005). Deconstructing, reconstructing, preserving Paul E. Meehl's legacy of construct validity. *Psychological Assessment, 17*, 415-422.

Manseau, H., Fernet, M., Hébert, M., Collin-Vézina, D., & Blais, M. (2008). Risk factors for dating violence among teenage girls under child protective services. *International Journal of Social Welfare, 17*, 236-242.

McCloskey, L., & Lichter, E. (2003). The contribution of marital violence to adolescent aggression across different relationships. *Journal of Interpersonal Violence, 18*, 390-412.

Mercer, S. L. (1987). Not a pretty picture: An exploratory study of violence against women in high school dating relationships. *Resources for Feminist Research, 17*, 15-25.

Miller, S., Williams, J., Cutbush, S., Gibbs, D., Clinton-Sherrod, M., & Jones, S. (2013). Dating violence, bullying, and sexual harassment: Longitudinal profiles and transitions over time. *Journal of Youth and Adolescence, 42*, 607-618.

Muñoz-Rivas, M. J., Gámez-Guadix, M., Fernández-González, L., & Lozano, M. P. G. (2011). Validation of the Attitudes about Aggression in Dating Situations (AADS) and the Justification of Verbal/Coercive Tactics Scale (JVCT) in Spanish adolescents. *Journal of Family Violence, 26*, 575-584.

Muñoz-Rivas, M. J., Graña, J. L., O'Leary, K. D., & González, M. P. (2007). Aggression in adolescent dating relationships: Prevalence, justification, and health consequences. *Journal of Adolescent Health, 40*, 298-304.

Muñoz-Rivas, M. J., Rodríguez, J. M. A., Gómez, J. L. G., O'Leary, D. K., & González, M. P. (2007). Valadación de la version modificada de la Conflict Tactics Scale (M-CTS) en población juvenile española. *Psicothema, 19*, 693-398.

Murphy, K. A., & Smith, D. I. (2010). Adolescent girls' responses to warning signs of abuse in romantic relationships: Implications for youth-targeted relationship violence prevention. *Journal of Interpersonal Violence, 25*, 626-647.

Murphy, K. A., Smith, D. I., & Xenos, S. (2012). TREAD: A promising change-target for partner

abuse prevention with adolescents. *Journal of Family Violence, 27*, 345-356.

National Institute of Justice. (2013, February). *Research and evaluation on violence against women: Teen dating violence, sexual violence, and intimate partner violence – FY 2013*. Retrieved from https://ncjrs.gov/pdffiles1/nij/sl001050.pdf

Nocentini, A., Menesini, E., Pastorelli, C., Connolly, J., Pepler, D., & Craig, W. (2011). Physical dating aggression in adolescence: Cultural and gender invariance. *European Psychologist, 16*, 278-287.

Nunnally, J. C. (1967). *Psychometric theory.* New York, NY: McGraw-Hill.

Offenhauer, P., & Buchalter, A. (2011). *Teen dating violence: A literature review and annotated bibliography* (DOJ Document No. 235368). Washington, DC: Library of Congress.

Offer, D., Kaiz, M., Howard, K. I., & Bennett, E. S. (2000). The altering of reported experiences. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 735-742.

O'Leary, K. D., & Slep, A. M. S. (2003). A dyadic longitudinal model of adolescent dating aggression. *Journal of Clinical Child and Adolescent Psychology, 32*, 314-327.

O'Leary, K. D., Slep, A. M. S., Avery-Leaf, S., & Cascardi, M. (2008). Gender differences in dating aggression among multiethnic high school students. *Journal of Adolescent Health, 42*, 473-479.

Orlando, M., Jaycox, L. H., McCaffrey, D. F., & Marshall, G. N. (2006). *Improving the measurement of socially unacceptable attitudes and behaviors with item response theory* (RAND Health WR No. 383). Santa Monica, CA: RAND Health.

Orpinas, P., Hsieh, H., Song, X., Holland, K., & Nahapetyan, L. (2013). Trajectories of physical dating violence from middle to high school: Association with relationship quality and acceptability of aggression. *Journal of Youth and Adolescence, 42*, 551-565.

Pan, H. S., Neidig, P. H., & O'Leary, K. D. (1994). Male-female and aggressor-victim

differences in the factor structure of the Modified Conflict Tactics Scale. *Journal of Interpersonal Violence, 9*, 366-382.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Paulhus, D. L. (1991). Balanced inventory of desirable responding. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 37-41). San Diego, CA: Academic Press.

Payne, D. L., Lonsway, K. A., & Fitzgerald, L. F. (1999). Rape myth acceptance: Exploration of its structure and its measurement using the Illinois Rape Myth Acceptance Scale. *Journal of Research in Personality, 33*, 27-68.

Pradubmook-Sherer, P. (2009). Prevalence and correlates of adolescent dating violence in Bangkok, Thailand. *Journal of Sociology and Social Welfare, 36,* 9-37.

Pradubmook-Sherer, P. (2011). Youth attitudes toward dating violence in Thailand. *International Journal of Offender Therapy and Comparative Criminology, 55*, 182-206.

Pradubmook-Sherer, P., & Sherer, M. (2011). Attitudes toward dating violence among Israeli and Thai youth. *Journal of Social and Personal Relationships, 28*, 809-828.

Price, E. L. (2000). *Risk factors for boys' psychologically abusive behaviour in dating relationships.* (Unpublished doctoral dissertation). University of New Brunswick, Saint John, New Brunswick.

Price, E. L., Byers, E. S., and the Dating Violence Research Team. (1999). The Attitudes Towards Dating Violence Scales: Development and initial validation. *Journal of Family Violence, 14*, 351-375.

Protivnak, J. J., & McRoberts, J. L. (2011). Abusive partner relationships in secondary schools: Identification and intervention strategies for school counselors. *Australian Journal of*

Guidance and Counselling, 21, 49-59.

Pyle, K. W. (1990). *Adolescent precursors of adult battered women: A postdictive study.* (Unpublished doctoral dissertation). Wright Institute Graduate School of Psychology, Berkeley, CA.

Rathus, J. H., & Feindler, E. L. (2004). *Assessment of partner violence: A handbook for researchers and practitioners.* Washington, DC: American Psychological Association.

Rayburn, N. R., Jaycox, L. H., McCaffrey, D. F., Ulloa, E. C., Zander-Cotugno, M., Marshall, G. N., & Shelley, G. A. (2007). Reactions to dating violence among Latino teenagers: An experiment utilizing the Articulated Thoughts in Simulated Situations paradigm. *Journal of Adolescence, 30*, 893-915.

Reyes, H. L. M., & Foshee, V. A. (2013). Sexual dating aggression across grades 8 through 12: Timing and predictors of onset. *Journal of Youth and Adolescence, 42*, 581-595.

Richards, T. N., & Branch, K. A. (2012). The relationship between social support and adolescent dating violence: A comparison across genders. *Journal of Interpersonal Violence, 27*, 1540-1561.

Roberts, T. A., Klein, J. D., & Fisher, S. (2003). Longitudinal effect of intimate partner abuse on high-risk behavior among adolescents. *Archives of Pediatrics and Adolescent Medicine, 157,* 880-881.

Roscoe, B., & Callahan, J. E. (1985). Adolescents' self-report of violence in families and dating relations. *Adolescence, 20*, 545-553.

Roscoe, B., & Kelsey, T. (1986). Dating violence among high school students. *Psychology, 23*, 53-59.

Schiff, M., & Zeira, A. (2005). Dating violence and sexual risk behaviors in a sample of at-risk Israeli youth. *Child Abuse & Neglect, 29*, 1249-1263.

Schultz, D. J., & Jaycox, L. H. (2008). Fear in adolescent dating relationships. *Journal of Aggression, Maltreatment & Trauma, 17*, 245-261.

Schumacher, J. A., & Slep, A. M. S. (2004). Attitudes and dating aggression: A cognitive dissonance approach. *Prevention Science, 5*, 231-243.

Scientific Advisory Committee of Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*, 193-205.

Sears, H. A., & Byers, E. S. (2010). Adolescent girls' and boys' experiences of psychologically, physically, and sexually aggressive behaviors in their dating relationships: Co-occurrence and emotional reaction. *Journal of Aggression, Maltreatment & Trauma, 19*, 517-539.

Sears, H. A., Byers, E. S., & Price, E. L. (2007). The co-occurrence of adolescent boys' and girls' use of psychologically, physically, and sexually abusive behaviours in their dating relationships. *Journal of Adolescence, 30*, 487-504.

Seidlitz, L., & Diener, E. (1998). Sex differences in recall of affective experiences. *Journal of Personality and Social Psychology, 74*, 262-271.

Sherer, M. (2009). The nature and correlates of dating violence among Jewish and Arab youths in Israel. *Journal of Family Violence, 24*, 11-26.

Sherer, M. (2010). Attitudes toward dating violence among Jewish and Arab youth in Israel. *Youth & Society, 42*, 132-150.

Sherer, P. (Pradabmook), & Sherer, M. (2008). Exploring reciprocity in dating violence among Jewish and Arab youths in Israel. *International Journal of Intercultural Relations, 32*, 17-33.

Shields, A., Gwaltney, C., Paty, J., & Shiffman, S. (2006). *Documentation of PRO instruments to meet contemporary FDA standards*. Retrieved from http://www.invivodata.com/media/files/PRO%20Consulting_White%20Paper_Documentatio

n%20of%20PRO%20Instruments.pdf

Slep, A. M. S., Cascardi, M., Avery-Leaf, S., & O'Leary, K. D. (2001). Two new measures of attitudes about the acceptability of teen dating aggression. *Psychological Assessment, 13*, 306-318.

Smith, G. A. (2005a). On the complexity of quantifying construct validity. *Psychological Assessment, 17*, 413-414.

Smith, G. A. (2005b) On construct validity: Issues of method and measurement. *Psychological Assessment, 17*, 396-408.

Smith, P. H., Earp, J. A., & DeVellis, R. (1995). Measuring battering: Development of the Women's Experience with Battering (WEB) scale. *Women's Health*: *Research on Gender, Behavior, and Policy, 1*, 273-288.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) scales. *Journal of Marriage and Family, 41,* 75-88.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283-316.

Swart, L., Seedat, M., Stevens, G., & Ricardo, I. (2002). Violence in adolescents' romantic relationships: Findings from a survey amongst school-going youth in a South African community. *Journal of Adolescence, 25*, 385-395.

Taylor, S. E. (1992). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*, 67-85.

Teitelman, A. M., Ratcliffe, S. J., Dichter, M. E., & Sullivan, C. M. (2008). Recent and past intimate partner abuse and HIV risk among young women. *Journal of Obstetric, Gynecologic & Neonatal Nursing, 37*, 219-227.

Teten, A. L., Ball, B., Valle, L. A., Noonan, R., & Rosenbluth, B. (2009). Considerations for the definition, measurement, consequences, and prevention of dating violence victimization among adolescent girls. *Journal of Women's Health, 18*, 924-927.

Thompson, M. P., Basile, K. C., Hertz, M. F., & Sitterle, D. (2006). *Measuring intimate partner violence victimization and perpetration: A compendium of assessment tools*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Ulloa, E. C., Jaycox, L. H., Marshall, G. N., & Collins, R. L. (2004). Acculturation, gender stereotypes, and attitudes about dating violence among Latino youth. *Violence and Victims, 19*, 273-287.

Vagi, K. J., Rothman, E. F., Latzman, N. E., Tharp, A. T., Hall, D. M., & Breiding, M. J. (2013). Beyond correlates: A review of risk and protective factors for adolescent dating violence perpetration. *Journal of Youth and Adolescence, 42*, 633-649.

Vézina, J., Hébert, M., Poulin, F., Lavoie, F., Vitaro, F., & Tremblay, R. E. (2011). Risky lifestyle as a mediator of the relationship between deviant peer affiliation and dating violence victimization among adolescent girls. *Journal of Youth and Adolescence, 40*, 814-824.

Walton, M. A., Cunningham, R. M., Goldstein, A. L., Chermack, S. T., Zimmerman, M. A., Bingham, C. R., . . . Blow, F. C. (2009). Rates and correlates of violent behaviors among adolescents treated in an urban emergency department. *Journal of Adolescent Health, 45*, 77-83.

Wekerle, C., & Tanaka, M. (2010). Adolescent dating violence research and violence prevention: An opportunity to support health outcomes. *Journal of Aggression, Maltreatment & Trauma, 19*, 681-698.

Wolfe, D. A., Scott, K., Reitzel-Jaffe, D., Wekerle, C., Grasley, C., & Straatman, A. (2001).

Development and validation of the Conflict in Adolescent dating Relationships Inventory. *Psychological Assessment, 13*, 277-293.

Wolfe, D. A., Wekerle, C., Scott, K., Straatman, A. L., Grasley, C., & Reitzel-Jaffe, D. (2003). Dating violence prevention with at-risk youth: A controlled outcome evaluation. *Journal of Consulting and Clinical Psychology, 71*, 279-291.

Wolitzky-Taylor, K. B., Ruggiero, K. J., Danielson, C. K., Resnick, H. S., Hanson, R. F., Smith, D. W., . . . Kilpatrick, D. G. (2008). Prevalence and correlates of dating violence in a national sample of adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*, 755-762.

Zaza, S. Wright-De Agüero, L. K., Briss, P. A., Truman, B. I., Hopkins, D. I., Hennessy, M. H., …Task Force on Community Preventive Services. (2000). Data collection instrument and procedure for systematic reviews in the *Guide to Community Preventive Services*. *American Journal of Preventive Medicine, 18*, 44-74.

CHAPTER FIVE

DEVELOPMENT OF A MEASURE OF PSYCHOLOGICAL AGGRESSION FOR TEEN DATING RELATIONSHIPS

Development of a Measure of Psychological Aggression for Teen Dating Relationships

Deinera Exner-Cortens

John Eckenrode

Dawn Schrader

Emily Rothman

Author Note

Deinera Exner-Cortens, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

John Eckenrode, Department of Human Development and Bronfenbrenner Center for Translational Research, Cornell University.

Dawn Schrader, Department of Communication, Cornell University.

Emily Rothman, Department of Community Health Sciences, Boston University School of Public Health.

This research was supported in part by a Doctoral Foreign Study Award from the Canadian Institutes of Health Research and a doctoral dissertation grant from the Department of

Human Development, Cornell University. The data reported in this paper were previously presented at the 121st Annual Convention of the American Psychological Association.

Correspondence concerning this article should be addressed to Deinera Exner-Cortens, CAMH Centre for Prevention Science, 100 Collip Circle, Suite 100, London, ON N6G-4X8. Contact: dme56@cornell.edu

Abstract

Psychological aggression in intimate relationships is a difficult construct to measure, due to the complex and nuanced nature of interpersonal interactions. While some work focusing on the improved measurement of this construct has occurred in adult samples, less work has occurred in adolescent populations. However, adolescent romantic relationships are an important developmental experience, and empirical data suggests that psychological aggression is common in teen dating relationships, leading to questions about the impact of this aggression on healthy development. Before these questions can be adequately answered, measures must be designed that capture psychological aggression that is purposeful, serious, and perceived as harmful. This paper reports on the initial adaptation of a measure of severe psychological aggression (the Measure of Psychologically Abusive Behaviors; Follingstad, 2011) for teen dating relationships. Adaptations were made based on the recommendations of 10 adolescent focus groups. A preliminary measure is presented, and future directions for the quantitative testing of this measure are described.

Development of a Measure of Psychological Aggression for Teen Dating Relationships

"I honestly think mental and emotional abuse is more, more uh, harmful or hurtful than physical. It's just my outlook on it because you can recover from physical but you'll always have that mental and emotional abuse. You'll always have it, it'll stay with you. It can't just go away."

- Adolescent Focus Group Participant (Rural, Senior High Girl)

In her statement, this teenage research participant captures a commonly expressed and empirically investigated idea about psychological abuse: that the experience of psychological abuse is as bad, or worse, than physical abuse. On the message boards of Pandora's Aquarium, a website for survivors of rape and sexual assault, there is a thread dedicated to psychological abuse and domestic violence, where the author of an essay on this topic states that "psychological abuse…[is] deeply, deeply wounding: a slap can seem like a loss of control in a moment of anger, but the psychological abuse is so personal and so insulting. When someone knows us well, they know our weakest and more vulnerable points, they know which parts of our bodies we hate, what our dark secrets are, and what certain words and phrases can tear us apart" (Sadie, 2010). Practitioners who work with victims and survivors often hear similar stories, where work with individuals who have experienced psychological abuse can be especially daunting, because the physical injuries may have healed, while the emotional ones have not, and may not for a long time.

On the empirical side, studies report that psychological abuse is often as harmful, or more harmful, than physical abuse for both men and women (O'Leary, 1996; Coker et al., 2002). In one of the first (and often cited) studies of this phenomena, Follingstad, Rutledge, Berg, Hause and Polek (1990) found that 72% (n=159) of the women in their study, all of whom had experienced some physical abuse, reported that received psychological abuse had a more severe impact on them than received physical abuse (there was no difference in severity or frequency of

physical abuse among women who did and did not find psychological abuse more severe). Since this study, several others have demonstrated similar findings. For example, in their study of domestic violence victims, Sackett and Saunders (1999) found that psychological abuse accounted for a much larger proportion of variance in a model predicting fear than did physical abuse. Perhaps not surprisingly, then, psychological abuse is not only related to adverse health and well-being in empirical work, but has also been found to be more strongly related to some negative outcomes than the experience of physical abuse. In a study of Italian women in a shelter, Baldry (2003) found that self-report psychological abuse was related to anxiety/depression, low self-esteem, and PTSD intrusion/avoidance symptoms, even when controlling for physical violence, which was not significantly associated with any of these symptoms. Lawrence, Yoon, Langer and Ro (2009) found that psychological, and not physical, aggression was longitudinally related to depression and anxiety in their community sample of newly married couples; in this sample, approximately 80% of husbands and wives reported some use of psychological aggression against the other partner (measured using a self-report checklist). And, using a nationally representative sample of women, Basile, Arias, Desai and Thompson (2004) found that self-report psychological abuse (assessed as power and control tactics and threats to harm or kill) was significantly associated with PTSD symptoms among those who had experienced violence from a current partner in the last year, controlling for physical violence and stalking.

Definitional Issues

These research findings, as well as practice-based knowledge and experience, have culminated in great concern over psychological abuse. However, despite this body of evidence, it is not always clear what studies are assessing when they consider 'psychological abuse.' For example, is it one construct, or several? Work with men in a batterer intervention program

suggests that different forms of psychological abuse (i.e., verbal abuse, controlling behaviors, and threats, as described by their partners) have different risk factors, and that these different forms may not represent an inclusive class or continuum of behaviors (Gondolf, Heckert, & Kimmel, 2002). In work with victims, Marshall (1996) found six distinct clusters of experiences with psychological abuse, based on frequency and type. Different types of psychological abuse (e.g., dominating/isolating vs. emotional/verbal) have also been found to have differential associations with depression in a sample of college women (Katz & Arias, 1999) and in a community sample of newly married couples (Lawrence et al., 2009), and Outlaw (2009) reported differential associations between different types of self-report psychological abuse (emotional, social, and economic) and the experience and frequency of physical abuse. These complicated findings suggest that current measurement of psychological abuse as a singular, behavioral construct may fail to capture important nuances.

Measurement of Psychological Abuse

O'Leary (1999) argues that psychological aggression in adults is measured reliably, and that commonly used measures have evidence of construct validity; however, in this same article, he states that "adequate definitions of psychological abuse in relationships do not exist for legal and formative diagnostic purposes" (p. 18). If we cannot provide adequate definitions for these purposes, do adequate definitions exist for research purposes? In a consideration of this question, Follingstad (2007) (the author of the article that began the previously described line of inquiry; Follingstad et al., 1990), offers a critique of the psychological aggression literature, and concludes that we do not have an adequate definition.[21] In her article, she covers four issues in this literature, the first of which is that psychological abuse "has not been adequately defined,

[21] Even before Follingstad's (2007) critique, deficiencies in the assessment of psychological aggression were noted. For example, in their review of the dating violence literature in the early 1990s, Sugarman & Hotaling (1991) stated that they did not include psychological abuse because "although recognized as a means to control another person, not enough work focused on operationalizing psychological abuse in the dating violence research" existed to justify its inclusion (p. 101).

validated, and conceptually anchored in a way which provides a sound basis for its measurement or which allows for conclusions to be made" (p. 441). Providing a review of the varying research bases used to define psychological abuse, and the incongruities among them, Follingstad (2007) highlights the troubled definitional state of the field, and provides support for her three other arguments, namely, the issues arising from defining psychological aggression parallel to physical abuse; the limitations of defining psychological aggression solely by self-report, without a consideration of the contexts of interpersonal interactions; and the inability of current measures to capture the complex and nuanced experience of psychological aggression (p. 441). In sum, "major shifts are required to advance the field, such as viewing human nature in a more mature and savvy fashion, recognizing that psychological aggression rarely occurs in a vacuum, and acknowledging that psychological 'abuse' may never be measured through pencil/paper formats" (p. 455).

Follingstad (2007) also advocates for the use of the term psychological aggression, instead of the term psychological abuse, because psychological aggression "covers a *range* of behavior, does not require a threshold severity level, and can consider whether an impact occurred, but does not have to require that a person has been harmed" (p. 443). Taking this definition further, Jordan, Campbell and Follingstad (2010) describe psychological aggression as "rang[ing] from boorish and inept relationship behavior through interpersonal terrorism," and in reference to thresholds, clarify that "there is currently no gold standard for when the threshold of 'psychological abuse' has been reached in a particular relationship" (p. 610), making the use of the term psychological aggression, which does not imply a threshold, more appropriate than the term psychological abuse. However, it is important to note that even if all these "boorish" behaviors could be identified, understanding when the "threshold of psychological abuse" has been reached also requires knowledge about intent, perception and capacity of the behavior to

harm the recipient (Jordan et al., 2010), and so a list alone would not fully elucidate the construct.

In a subsequent review article, Follingstad (2009) argues that due to these methodological and measurement issues, the research findings discussed previously should be interpreted with caution, and that the most consistent finding is that psychological aggression is related to leaving (or desiring to leave) the intimate relationship, and not any particular adverse outcome (although there is also relatively good support for associations with depressive symptomatology). As discussed by Jordan et al. (2010), "although many articles report that deleterious effects have been identified as a result of being the target of psychological aggression, the subjective nature of the experience, the occurrence of it within interpersonal interactions and relationship contexts, the frequent co-occurrence of physical violence, and problematic measurement all combine to reduce certainty as to the impact of these actions" (p. 614).

Thus, the field appears to be dominated by measures that inadequately capture the phenomena of interest (psychological abuse); rather they capture some form of psychological aggression, much of which may be potentially normative and not harmful over the long term. Against this, however, is the practice-based knowledge of the harm caused by psychologically abusive behaviors, a knowledge which we fail to represent using the current modes of assessment. Thus, the measurement of psychological aggression is (or should be) an active and important area of work.

Adolescents and Psychological Aggression

All the data and arguments thus far presented in this paper pertain to adult domestic violence; however, as discussed in other papers in this volume, adolescents also use and experience psychologically aggressive actions within the context of their dating relationships. Follingstad (2009) seems somewhat dismissive of these actions, stating "studies on dating

populations, with often short-term relationships, are probably less important to consider than studies using couples in long-term cohabitating relationships because of the low level of psychological aggression or depression" (p. 277, in reference to understanding associations between psychological aggression and outcomes). However, neither of these latter statements appear to be the case. Using nationally representative data, Kessler et al. (2010) found that 10.2% of adult women and 6.2% of adult men had a major depressive episode within the past year, compared to 14.0-21.5% of female adolescents and 4.4-11.0% of male adolescents aged 15-18 (Kessler & Walters, 1998). And, psychologically aggressive actions are common in adolescent romantic relationships (approximately 30% lifetime prevalence in a nationally representative sample; Halpern, Oslak, Young, Martin, & Kupper, 2001), just as they are in adult relationships (where national estimates of lifetime prevalence are around 50%; Black et al., 2011). Emerging research also demonstrates longitudinal associations between experiences of psychological aggression and multiple adverse health outcomes (Exner-Cortens et al., 2013), just as research with adults has shown negative consequences. In regards to perception, Jouriles, Garrido, and Rosenfield (2009) found that in their sample of high school students (mean age=15.4), those who had experienced both physical and self-report psychological aggression rated psychological aggression as more unpleasant than physical aggression, and were less likely to ascribe psychological aggression to a partner playing around, seeming to reflect findings from samples of adult women. Finally, adolescent romantic relationships are important developmental experiences (Collins, 2003; Feiring & Furman, 2000; Furman & Shaffer, 2003), and the topic of a growing research literature, evidence that these relationships are no longer considered trivial (Collins, Welsh, & Furman, 2009).

Thus, it appears that psychological aggression in adolescent romantic relationships is equally deserving of proper measurement; however, the deficiencies outlined by Follingstad

(2007) regarding the measurement of adult psychological aggression also pertain to adolescent measurement, with the additional caveat that many of the items used for adolescents were originally developed for adult samples, often with little to no rigorous adaptation work (see Exner-Cortens, Gill, & Eckenrode, Chapters Three and Four). This work is important though, since the interpretation of behaviors depends on context, including the context of the relationship itself (Murphy & Cascardi, 1993), and adolescent romantic relationships are certainly a very different context (developmentally and otherwise) than married or long-term adult relationships. In light of this, Follingstad (2009) recommends that assessments of psychological aggression in dating relationships be done separately from those in married or long-term cohabitating relationships, in part because psychological aggression experienced in short-term dating relationships may often be normative. This seems a prudent suggestion, given that adolescents are learning to negotiate conflict in intimate relationships characterized by heightened emotionality and volatility (Feiring & Furman, 2000; Giordano, Manning, & Longmore, 2006; Wolfe & Feiring, 2000), and so the function of some of this aggression may differ from adult relationships. For example, in their validation of the Conflict in Adolescent Dating Relationships Inventory (CADRI), Wolfe et al. (2001) found that use of relationally aggressive behavior (e.g., spreading rumors) was most strongly related to a latent 'abuse' factor among grade 9 students, compared to students in grades 10 and 11, suggesting that the use of these behaviors by some younger adolescents may be indicative of a developmental process, and not negative intent. However, even given this knowledge, how best to assess this aggression remains an active question.

New Directions in Assessment

Arguments that there are important and nuanced differences in experiences of psychological aggression, experiences that are not captured with simple checklists or frequency

counts, are persuasive; how, then, do researchers begin to differentiate normative psychological aggression (i.e., aggression which is often used in relationships, and not harmful beyond hurt feelings; Follingstad & Rogers, 2013) from the purposeful, harmful aggression that, as described by Sackett and Saunders (1999), "if severe enough…may lead to self-doubt, confusion, and depression" (p. 105)? In short, how severe is severe *enough*, and how do we begin to assess this?

Follingstad and Rogers (2013) suggest that to conduct this type of assessment, "measurement must capture serious actions, malignant intent, a range of psychological behaviors broad enough to cover the domain of harmful actions, and some sense that the context supports labeling the action as abuse" (p. 12). To incorporate this definition into measurement, Follingstad (2011) created a measure of severe psychological aggression (Measure of Psychologically Abusive Behaviors; MPAB), drawing upon a review of the psychological aggression literature as well as her own previous work. The 42-item MPAB contains mild, moderate and severe items representing 14 categories of psychological aggression; each item also contains intention of the actor in its description of the action. For example, under the category *manipulation*, the three actions are "continued to act very upset (e.g., pouted, stayed angry, gave you the silent treatment) until you did what he/she wanted you to do" (*mild*); "threatened to end the relationship as a way to get you to do what he/she wanted" (*moderate*); and "threatened to commit suicide as a way to get you to do what he/she wanted" (*severe*). Compared to Follingstad's (2005) prior measure of psychologically aggressive actions (the Follingstad Psychological Aggression Scale, FPAS), which considered the full range of these behaviors (including items that might represent more normative interactions), items on the MPAB were rated as more severe by a nationally representative sample of adults: on the FPAS, the mean severity (as rated by college students) indicated that respondents felt the actions were moderately aggressive, while on the MPAB, the mean severity indicated that respondents felt these actions were strong violations of reasonable

relationship behavior (Follingstad et al., 2005; Follingstad, 2011). To assess other contextual variables, including perception, Follingstad and Edmundson (2010) created MPAB follow-up prompts. If a respondent endorsed any of the MPAB items, they were prompted to report on the frequency of the behavior, the extent to which the behavior impacted them emotionally, the extent to which the behavior impacted them behaviorally, and the extent to which they considered their partner to have been psychologically abusive.

Though the MPAB is a potentially promising measure to assess severe psychological aggression in adolescent romantic relationships as well, it has so far only been used in adult samples. It is thus possible that all included items are not seen by adolescents as especially severe, or that items viewed as severe by adolescents are not included in the adult measure. Additionally, the MPAB contains no items on electronic psychological aggression (i.e., psychologically aggressive actions perpetrated through social media, email, instant messaging or cell phones), though emerging evidence suggests that this is an important form of psychological aggression among teens (Draucker & Martsolf, 2008; Offenhauer & Buchalter, 2011). Therefore, prior to being used to assess psychological aggression in adolescent romantic relationships, the MPAB needs to be adapted and then validated for use in teen populations. The purpose of this study was to perform focus groups in order to adapt the MPAB for use in early and mid-adolescent samples, and to add questions on electronic aggression, with the goal of producing an initial measure, the Psychological Aggression in Teens Scale (PATS), for quantitative testing.

Method

Participants

Recruitment locations from across New York State were contacted with the assistance of the Bronfenbrenner Center for Translational Research at Cornell University. These locations all participated in on-going research pertaining to sexual health education evaluation. Sites were

chosen to represent the geographic distribution of New York State (upstate urban, rural and downstate). Participants attended programs at each of the chosen locations, and were recruited based on their interest and availability. To be eligible to participate, participants needed to be enrolled in middle school, high school or a GED program (and/or be age 18 or younger) and provide signed parental consent. In total, we recruited 64 participants (48.4% female). Other demographic data are listed in Table 5.1. At each location, we conducted two groups (1 male, 1 female), for a total of 10 groups.

Based on the recommendations of Krueger and Casey (2009), we restricted groups to a maximum of ten people. We also kept groups age-homogenous (i.e., either all senior high or all middle school participants). While we had planned to conduct an equal number of senior high and middle school groups, middle school participants were difficult to recruit, and so eight of the 10 groups were conducted with high school aged participants. Focus groups were approximately 60 minutes in length, and were completed in Fall 2012. This study was reviewed and approved by the Cornell University Institutional Review Board.

Table 5.1
Demographics

Age, y, mean (SD); range[a]	14.82 (1.94); 11-18
	% (n)
Sex	
Male	51.6 (33)
Female	48.4 (31)
Race/ethnicity	
White	15.6 (10)
Black	39.1 (25)
Hispanic	17.2 (11)
Multiracial	25.0 (16)
Other	3.1 (2)
Family structure	
Two biological parents	34.4 (22)
One biological parent and one step-parent	18.8 (12)
Single parent (biological)	39.1 (25)
Other	7.8 (5)

(Table 5.1 continues)

(Table 5.1 continued)

	% (n)
Parental education[b]	
Less than high school	14.1 (9)
High school graduate	10.9 (7)
Some college education	20.3 (13)
College graduate or more	39.1 (25)
Grade in school[c]	
6th	6.3 (4)
7th	15.6 (10)
8th	6.3 (4)
9th	17.2 (11)
10th	14.1 (9)
11th	25.0 (16)
12th/GED	14.1 (9)
Past year dating violence victimization (% yes)[d]	
Sworn at	42.1 (24)
Insulted, called names or treated disrespectfully in front of others	21.1 (12)
Threatened with violence	7.0 (4)
Hit, slapped or physically hurt on purpose	12.3 (7)

[a]*n*=62
[b]*n*=54
[c]*n*=63
[d]Of those who had dated (*n*=57).

Procedure

At the beginning of each group, the purpose of the focus group was explained to the participants, and parental consent forms were collected. Participants were also asked to provide their own assent, and were told that they could leave the group at any time. No participants refused to sign the assent or left the group. We gave each participant $15 as a thank-you for participation.

Participants were then asked to fill out a short warm-up questionnaire, to familiarize them with the focus group topic. We also collected this data due to the potential for group influence during the focus group stickering task (see below). The warm-up questionnaire contained a random selection of 14 of the 42 MPAB items (mild, moderate, severe), and one of three investigator-added questions ("kept you from getting the medical care you needed," "did not

want you to go to school or other social activities," "restricted your use of the telephone or internet"). Participants were asked to indicate whether the action was something someone their age might do to harm a dating partner (*yes*, *no* or *unsure*), and if this action was a violation of what should happen in a dating relationship (where 1=*not a violation at all*, 10=*worst possible violation*).

Participants then reviewed five posters, containing the 42 MPAB items and the three investigator-added questions. They were instructed to read each item, and then use stickers to indicate whether they thought the action was something someone their age might use to harm a dating partner: a green sticker indicated that they thought it was something someone their age would do, a red sticker indicated that it was not, and a blue sticker indicated that they didn't know or were unsure. They were also told to place their stickers based on their own thoughts and opinions, and not the thoughts and opinions of others in the group. For both the warm-up questionnaire and sticker activity, a dating relationship was defined as *the kind of relationships you and your close friends experience, where you like a person, they like you back, and other people know that you are together*.

The sticker patterns guided the subsequent focus group discussion. Questions that were specifically focused on were those with a lot of red stickers (indicating that many participants thought the item was not applicable for their age group), a lot of blue stickers (indicating that the participants were unsure about the item), and items with very mixed sticker patterns (indicating a lot of disagreement). Participants were asked to discuss why they had selected a particular sticker color, in order to identify needed question revisions, deletions and additions. At the end of the discussion, participants were asked to describe ways that technology could be used to harm someone in a dating relationship, and to describe any other questions they thought we should add, remove or modify.

At the end of the focus group, all participants completed a brief demographics questionnaire. Questions included race/ethnicity, family structure, socio-economic status (as indicated by parental education), current grade in school, and information on social media use. Social connectedness was assessed with the question "I have a number of people (or a group) to interact and connect with, and I feel like I belong with them," using a 5-point Likert-type scale (1=*strongly disagree*, 5=*strongly agree*). Participants were also asked to describe their dating history (number of dating relationships and gender of partner(s)), and to indicate if they had experienced physical or psychological dating violence victimization in the past 12 months, using questions from the National Longitudinal Study of Adolescent Health (three psychological items) and the Youth Risk Behavior Surveillance Survey (one physical item).

All materials were piloted with a high-school aged male and female from an upstate urban location and with a group of adolescents in New York City prior to the focus groups. The first author [DE-C] moderated all focus groups.

Data Analysis

All focus groups were recorded, and each recording was transcribed verbatim by undergraduate research assistants who were supervised by the first author. All transcripts were coded in Atlas.Ti using directed content analysis (Hsieh & Shannon, 2005), also by undergraduate research assistants supervised by the first author. For both coding and transcription, all research assistants received multi-hour training from the first author, including performing practice transcription and coding on data from other projects. Practice work was reviewed during team meetings, and any issues resolved before research assistants were allowed to transcribe/code data from this project. When transcribing and coding focus group transcripts, research assistants worked in teams of two: one member of the team was assigned as the primary transcriber/coder, and the other team member served as a consecutive secondary

transcriber/coder. The secondary transcriber/coder reviewed the primary research assistant's work, and the two team members then met to discuss any differences and resolve discrepancies, until consensus was reached. All coded transcripts were also reviewed by the first author. The codebook was developed *a priori* by the research team, and updated in an iterative fashion during the coding process. Codes pertaining to the MPAB questions are the focus of the present article; all codes were reviewed by age/gender group (i.e., senior high girls, senior high boys, middle school boys and middle school girls).

Quantitative data from the posters (sticker counts), warm-up questionnaire and the demographic questionnaire were analyzed in SPSS V21.

Results

Demographics

We conducted ten groups, five male and five female, with a total of 64 participants. Four groups were conducted in upstate urban locations, four in rural locations, and two in a downstate location. Eighty-percent of the groups were conducted with high-school aged youth; the two middle school groups were conducted in a rural location. Groups ranged in size from three to nine people.

The participants in our sample were aged 11-18 (Table 5.1), and 93.8% (*n*=60) went to public school. When asked how connected they felt to their peer group, the mean (SD) was 3.89 (0.99), indicating a moderate level of social connectedness. The most commonly used social media site was Facebook (76.6%), followed by Twitter (45.3%) and Instagram (21.9%). Social media use was listed by 98.4% (*n*=63) of participants. Other demographics are listed in Table 5.1.

Of the 89.1% of participants who had dated (*n*=57), 12.3% had experienced any physical dating violence (17.2% of boys and 7.4% of girls), and 47.4% had experienced any

psychological dating violence (44.8% of boys and 50.0% of girls). The prevalence of physical dating violence in this group was slightly higher than the statewide prevalence in 2011 (10.3% in a representative sample of grade 9-12 students; CDC, 2012).[22] Statewide numbers are not available for psychological violence victimization; however, the prevalence of past year psychological violence victimization reported in our sample was higher than the lifetime prevalence reported in a nationally representative sample of youth in grades 8-12 (29%; Halpern et al., 2001). In our sample, all participants who had experienced physical violence had also experienced psychological violence (of those who had dated, 12.5% experienced both physical and psychological violence, 33.9% experienced psychological violence only, and 53.6% experienced no dating violence).

Violation of Dating Behavior: Comparison of Adults and Adolescents

On the warm-up questionnaire, participants were asked if they thought the particular behavior might be used by someone their age to harm a dating partner, and how much the behavior violated what should happen in a dating relationship. The latter question was also asked by Follingstad (2011) in an adult sample, and a comparison of the results for the present adolescent sample and Follingstad's (2011) adult sample are presented in Table 5.2 and Figure 5.1. Across all question types (i.e., mild, moderate and severe), adolescents thought items were more of a violation than adults (Figure 5.1). Looking at specific questions, items where adolescents did *not* find the behavior more of a violation than adults were: the severe item from the *sadistic* group (1C), the moderate and severe items from the *isolating* group (3B and 3C), the moderate item from the *verbal abuse* group (6B), the moderate item from the *wound through sexuality* group (7B), the mild item from the *treat as inferior* group (8A), the mild item from the *hostile environment* group (10A), the severe item from the *wound through fidelity* group (11A),

[22] However, the statewide prevalence does not appear to exclude non-daters. If we include both daters and non-daters in our estimate, the prevalence of physical violence victimization was 10.9% (15.6% of boys and 6.7% of girls), similar to the statewide prevalence.

the mild and severe items from the *jealousy* group (12A and 12C), the moderate item from the *withheld emotional/physical* group (13B), and the mild item from the *control personal decisions* group (14A) (totaling 28.6% of the 42 MPAB questions; specific question text is given in Appendix 5.A).

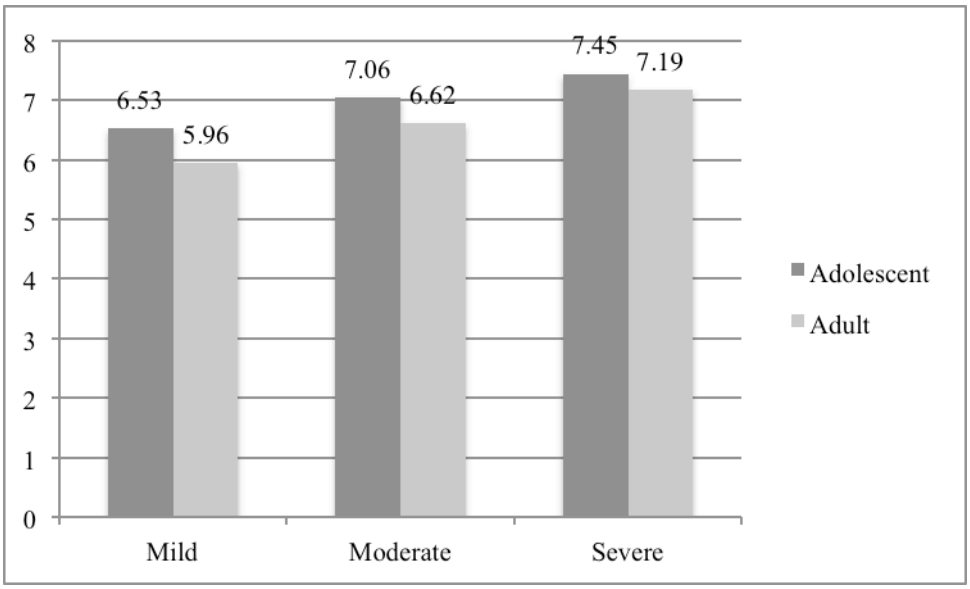

Figure 5.1. Comparison of average scores for adolescents and adults on mild, moderate and severe MPAB questions. Adult means were adapted from Follingstad (2011), Table 1.

Looking only at the adolescent results, a number of question groups did not follow the expected increasing violation pattern across mild, moderate and severe items (Table 5.2); however, it is important to note that these items were not seen by the same groups of adolescents (i.e., one group would see the mild item from a particular question grouping, another the moderate item, and another the severe item), and so participants could not compare across questions within a group to judge relative severity. The question groupings that do follow the expected pattern are *threats*, *manipulate*, *treat as inferior*, *hostile environment* and *control personal decisions* (five of 14 question groups). Failure to follow the expected pattern may also result from lack of relevance of a question to the adolescent age group; for example, the item "harmed pets as a way to intimidate you" (1C) was uniformly seen by participants in this sample as not applicable to teens (see below), which may explain its lower mean violation rating (Table

5.2). This is similar to the *isolating* group, where questions 3B and 3C were not consistently seen by participants as something a teen would do, and have a lower mean violation score than the mild item (3A), which was viewed as something a teen would do.

Table 5.2
MPAB Group Means in an Adolescent (n=64) and Adult (n=649) Sample

	Mild		**Moderate**		**Severe**	
	Adol.	***Adult***	***Adol.***	***Adult***	***Adol.***	***Adult***
Sadistic	8.36	7.73	8.81[a]	8.42	7.24[a]	8.46
Threats	7.86	5.81	8.67	8.25	9.52[a]	9.21
Isolating	7.67	6.67	6.82[a]	6.89	6.62[a]	7.30
Manipulate	6.36	5.44	7.00	6.64	8.86[a]	8.41
Public humiliation	6.95	6.74	7.91	6.82	7.00	6.89
Verbal abuse	8.18	6.51	6.57	6.82	7.57	6.82
Wound through sexuality	6.52	5.89	6.32[a]	6.76	8.77[a]	7.39
Treat as inferior	5.98	6.04	7.00	6.68	7.12[a]	6.93
Monitor	6.00	5.89	8.19	6.60	7.26[a]	6.82
Hostile environment	5.14	6.00	6.45[a]	6.23	7.52[a]	6.87
Wound through fidelity	6.91	5.79	6.52	6.10	6.76	6.82
Jealousy	4.95	5.32	7.80	5.60	6.68	6.97
Withheld emotional/physical	6.10[a]	4.88	5.14	5.65	6.86	5.77
Control personal decisions	4.50[a]	4.71	5.67[a]	5.24	6.57[a]	5.97
Overall mean	**6.53**	**5.96**	**7.06**	**6.62**	**7.45**	**7.19**

Note. N, adolescent, range 20-22. Adult means are from Follingstad (2011), Table 1, p. 1201. Abbreviations: Adol. – adolescent.
[a]Question requiring revision.

Questions Requiring Revision

To determine which MPAB questions were most in need of revision, the percentage of participants reporting that the item was something someone their age might do to harm a dating partner was reviewed for both the warm-up questionnaire and sticker data; if ≤ 50.0% of participants reported that an item was something someone their age would do (in the warm-up questionnaire, by circling no or unsure; in the sticker activity, by choosing a red or blue sticker), the item was considered as needing revision. In total, 19 questions met this criterion (16

questions from the original MPAB, and the 3 investigator-added questions). Six of these questions were identified as needing revision on both the warm-up questionnaire and sticker activity, ten on the sticker activity only, and three on the warm-up questionnaire only. Questions requiring revision are indicated in Appendix 5.A.

To determine required revisions, focus group discussions of each of the 19 identified questions were reviewed, and revisions made based on the ideas presented. For some questions, this led to the addition of several new questions. Upon reviewing these 19 questions, several dominant themes stood out as reasons for revision. For brevity, these common themes are reviewed below, as opposed to reviewing specific revisions for each individual question (for specific revisions, see Appendix 5.A).

Extreme. Ten of the nineteen questions were discussed as too extreme for teen dating relationships, by all groups of senior high girls (*n*=4), senior high boys (*n*=4) and middle school boys (*n*=1), but not the group of middle school girls. Senior high girls, senior high boys and middle school boys all used this reason to describe why questions 4C ("threatened to commit suicide as a way to get you to do what he/she wanted") and 10C ("followed or had you followed by someone else as a way of checking up on your activities") weren't generally relevant to teen dating relationships. For example, when discussing 4C, one senior high girl said,

> "I feel like also with teenage relationships you have, it's easier to get out of a relationship as a teenager, because you're not dependent on them for a house or for, finances, or anything like that so, leaving someone doesn't really…like yeah, it might affect you emotionally, but it's not going to affect like, eating, living, and like anything like that so I feel like if that, it would have been stopped before it got that far."
> *- Rural, Senior High Girls*

and a senior high boy from the same site said,

> "I just don't think that threatening to commit suicide is a good idea. Even a logical threat at all because it is a lose-lose situation. You are not getting anything out of it and neither is the significant other in the situation. And it just comes to be a level of maturity, who would even threaten to commit suicide over a relationship…I mean especially being a teenager."
> *- Rural, Senior High Boys*

Thus, both participants indicated that threatening to commit suicide was too extreme for their age group. However, not all participants felt this way. In a different group, a senior high girl stated,

> "So the little blue sticker that's mine. Because like I don't know I mean, it seemed pretty like drastic like something really seriously like 'Oh my god, I'm going to kill myself' but there is like some really dramatic people who do stuff like that."
> - *Upstate Urban 1, Senior High Girls*

indicating that while this action is extreme, it's possible that a "really dramatic" person might engage in this behavior.

Senior high boys and girls also thought questions 1C, 2C, 9C and one investigator-added question ("kept you from getting the medical care you needed") were too extreme. Of this group, question 1C was consistently viewed as especially extreme,

> "I put red…cuz what teenager threatens another teenager's pet?"
> - *Downstate, Senior High Girls*

> "You gotta be crazy."
> - *Upstate Urban 2, Senior High Boys, in reference to what he thought about 1C*

with participants having difficulty understanding why someone would harm a pet. As might be expected, the majority of the questions viewed as too extreme were Category C ('severe') questions; the only exceptions were 3B (senior high boys) and 1B (senior high girls).

Overall, then, participants in this group viewed a number of the 'severe' (Category C) items as too severe. However, it is important to note that the measure is intended to assess severe psychological aggression, and so viewing an item as extreme was not an automatic disqualifier. Rather, all comments about a particular question were reviewed, and appropriate revisions made (Appendix 5.A); this was possible for all questions but 1C ("harmed a pet as a way to intimidate you"), for which no appropriate revisions were suggested. Instead, this question was changed to "Threatened to hurt something that means a lot to you (for example, a pet) to intimidate you." In addition, two other questions from another measure of adolescent dating aggression (the APAS;

Leisen, 2000) that seemed to capture potentially sadistic behavior were added to category 1C, as these may prove more relevant in quantitative testing.

Developmental. This theme included discussion of how the behavior described by an item might apply to older or younger people, but not teenagers. Developmental reasons were discussed for 15 of 19 questions by all groups of senior high girls (*n*=4), senior high boys (*n*=4) and middle school girls (*n*=1); middle school boys did not discuss this reason. For example, in reference to question 14C ("tried to make major decisions that affected you without consulting you"), a high school female participant discussed a developmental reason when she said that "yeah, but like major decisions would like affect more older, so, like, having [an] example maybe would help" (*Upstate Urban 2*). In this quote, this participant also indicates a request that was common across questions – the addition of specific examples. To make questions more age-relevant, revisions involving specific examples were made to a number of questions, including 14C, where the suggested examples were college and birth control use (Appendix 5.A). Similarly, the item "ignored important holidays and events as a way to punish or hurt you" was initially viewed as not relevant (i.e., because "that's too much of an older person thing" (*Rural, Senior High Girls*)), but following discussion, participants felt it could be relevant with the inclusion of specific holidays, such as Valentine's Day and anniversaries.

Related to this theme were the ideas of a behavior being too advanced (i.e., more sophisticated than what a teenager would do; seven of 19 questions), or not applicable because of the duration of teen relationships (six of 19 questions). In regards to sophistication, a senior high girl said,

> "I put a blue sticker because like people my age, like the people I know and hear about they don't really have creative minds, they are more like the basic kind of people."
> *- Upstate Urban 2, Senior High Girls, in reference to 7C*

while in regards to duration, a senior high boy remarked,

> "Or, in the back of your mind those decisions that you can't approve of them or like most of the relationships when you're a teen you're not really thinking about being with that person for the rest of your life or like up to a year or two type of thing."
> *- Upstate Urban 2, Senior High Boys, in reference to 14C*

indicating that finding belittling ways to have sex (7C) was more complex than what an adolescent might do, and that the number of major decisions an adolescent could make for a dating partner (14C) was limited by the presumably short duration of their relationship. Similar comments were made about other questions covered by this theme, and adaptations were made to make questions less sophisticated and more relevant to the nature of teen dating relationships. For example, a revision to 7C removes the word belittling, which confused some participants, and instead states "made you do sexual things you were not comfortable with, in order to embarrass or humiliate you," which better captures how participants discussed this item (see also *Sex and Contraceptives*, below), and revisions to 14C include examples that participants suggested did apply to teen dating relationships (Appendix 5.A).

A number of groups (*n*=3 senior high girl groups, *n*=4 senior high boy groups and the middle school boys group) also discussed that a dating partner wouldn't be able to restrict behavior in the way described by the question, because it was not an adolescent dating partner's role, but rather, for example, the role of the parents. In reference to keeping a partner from accessing medical care, a senior high boy discussed that "you don't really own that person, they have parents that are gonna make them go get help" (*Upstate Urban 1*), while another said that "[my partner] can't take away my stuff. Well, if you my mom, well yeah, you can take away my stuff, {laughter} but you are not my mom, you can't take away my stuff, you can't tell me what to do…" (*Downstate*), in reference to whether a dating partner could restrict access to the telephone or internet. A girl from a rural site agreed, saying,

> "I mean family is really a big part of your life when you're a teenager, like your mom, or your dad, or whoever's taking care of you, takes, they're like your basis, that's where you go home, that's where you eat, that's what pays for everything for you, so really, like they can't control

anything that happens in that aspect."
- Rural, Senior High Girls, in reference to 3A

In this way, the supervisory role of parents was also viewed as protective against some forms of aggression that might be possible in an adult relationship, a belief shared by a number of participants in these groups. In addition to parental oversight, a dating partner may not be able to restrict behavior because, based on the original question, the scope seemed too broad,

"I would say no because our generation is like so full of technology that I don't even think it would be a point for me to say like to a girl 'I don't want you on Facebook.' Because there are so many ways for her to get on there without me knowing it and it would be stupid of me to tell her that because it's like her own life that's her own thing, you know what I'm saying? I couldn't see that happening."
- Upstate Urban 2, Senior High Boys, in reference to 10C

Thus, as with other questions in this category, revisions were made to make questions more specific. For example, in regard to the *monitoring* items (10A, 10B and 10C), specific examples of how social media could be used to facilitate monitoring were built into the question, as opposed to just referring to the Internet broadly.

Family vs. Friends. This theme was specific to questions that discussed both family and friends (e.g., 1B). For these questions, most participants felt that threats against an adolescent's family were not very realistic, but that threats against friends could happen. As discussed by one participant, "teen years, your friends are more a family than family" (*Rural, Senior High Girls*), and so the dating partner might have more access to friends than family, because "you see them [your partner] with their friends more than you see them with their family since you're in high school and things like that" (*Upstate Urban 1, Senior High Boys*). The two middle school groups also agreed that it would be more feasible to keep someone from seeing their friends than their family.

However, while it might be difficult for a dating partner to keep an adolescent from seeing their family (as described in 3A, 3B and 3C), this does not mean that threats against a

family member wouldn't be upsetting. As described by one senior high boy "if you're in a relationship, [and] you're tryin' to hurt what's closest to the person and what's closer than blood? So simple as that" (*Upstate Urban 1*, in reference to 1B). To involve family members, it might be possible to spread rumors, as described by one senior high girl,

> "R1: If they just make up like an outrageous lie about like my parent that's clearly a boundary and I don't play that.
>
> {M: Do you know what an outrageous lie would be?}
>
> R1: My mom is a slut or something. I don't know, cause obviously I mean that wouldn't be true and nobody wants anyone to talk…to talk badly about their parents, so."
> *- Upstate Urban 2, Senior High Girls, in reference to 3A*

In response to whether we should remove all family items, one senior high boy said,

> "It depends on the person, because I'm really more family oriented than friends and some other people care about their friends more."
> *- Downstate, Senior High Boys, in reference to 4B*

Thus, while the consensus was that it would be difficult to keep an adolescent from their family (as described in questions 3A-3C), and that it might be difficult to physically harm a family member (as described in question 1B), the idea of using family wasn't completely irrelevant. To reconcile these two ideas, the mention of family was removed from question 1B, 3A, 3B and 3C revisions, and an item on family was added to the question 5B revision ("Started rumors about your family in order to embarrass you").

Technology. The topic of technology (including social media and cell phones) was discussed by all groups, specifically in reference to an investigator-added question ("restricted your use of the telephone or internet"), and question 10C ("followed or had you followed by someone else as a way of checking up on your activities").[23] One group of senior high girls (*Upstate Urban 1*) also discussed the use of social media in regard to question 7C ("insisted you have sex with him/her in belittling or humiliating ways"), defining that belittling might be sex

[23] While use of technology was a question at the end of the focus group guide, in all groups, it was discussed before this specific question was asked.

that was "public or like on camera or something" and that this was something someone their age might do. These videos could also be used to manipulate the partner into having sex again, through threat of posting, a suggestion which was incorporated into one of the 7C revisions.

Regarding the investigator-added question and question 10C, both were seen as too broad as originally written. For the question about restricting telephone or Internet use, participants thought the words *Internet* and *telephone* were too general, as indicated in this conversation from a rural group,

> "R2: I feel like with the internet it'd be more social medias than just the internet...
>
> {everyone joins in agreement}
>
> R2: …and just social networking."
> - *Rural, Senior High Girls*

Participants also felt that, as worded, restricting Internet and telephone use didn't sound like something a dating partner could do,

> "It's just too controlling in a relationship nobody really gonna tell your partner 'oh you can't use your cell phone or internet' cuz first of all they not paying their partner's bills or anything so they won't be able to control what she do with her cell phone or have internet anyways."
> - *Upstate Urban 1, Senior High Boys*

As with many of the questions, then, this question could be improved by making it more specific and including examples. For example, instead of the use of the general word *Internet*, a participant described that a partner could try to control your social life, where social life is "everything – like who you talk to, what you're doing with you phone, what you do on the internet" (*Upstate Urban, Senior High Girls*). This definition was adopted into the question 3B revision. Participants also gave a number of specific examples about how email, cell phones and social media could be used to harm a dating partner:

> "Like you talking to somebody, probably talking to an older person or a family member and they mislead that like you're talking to an older person or something and they're saying, 'oh, I'm not enough for you or something?'"
> - *Downstate, Senior High Girls, in reference to how email could be used*

"Or even like the phone thing, like maybe they might see someone texting you, get jealous, or something like that, so take your phone or tell that person not to text you or call you anymore."
- Upstate Urban 2, Senior High Girls

"I've seen that done before when they come…I was in a relationship where every time I came over I had to hand my phone over like it was I couldn't talk, I couldn't text, even my mom called me. I'm telling you, it's crazy."
- Upstate Urban 2, Senior High Boys, in reference to a partner restricting phone use

"I put green because obviously, we're teenagers, we use our cell phones, that's our main connection to life…pretty much. Um your boyfriend at school, your boyfriend outside of school, just like you said, it's, don't use that when you're with me, this is my time. Internet, obviously, we're teenagers, we use Facebook, IM, 'what the hell are you doing on Facebook, I told you to go.' They can't restrict you from using the Internet. They could take your phone; they could smash your phone. Internet at home you get punished for. 'I'll leave you, get off Facebook. Delete your Twitter.'"
- Rural, Senior High Girls

The rich variety of examples given by participants in regards to how a partner could specifically restrict Internet and phone use were used to guide revisions, and the subsequent revised questions were incorporated into the *Isolating* question block as forms of electronic isolation (Appendix 5.A).

Possible revisions to question 10C were also discussed by a number of participants, since the original question didn't seem to fit the adolescent context,

"Yea, no one is going to waste their time like following someone around, especially teenagers. They got their own things going on."
- Rural, Senior High Boys

Suggested revisions to 10C fell into two broad categories – tracking using friends, and tracking using social media or other technology.[24] For example, friends could be used to subtly watch a dating partner, and report back on activities, as described by these participants:

"I was at a party, and like, my girl didn't go to the party but her friends were there, and then I heard a girl talking to her phone, I hear she was like 'yeah, we watching her' and I see her following me and I ain't getting no dubs [inaudible], so I'm like ai'ght, I left the party. But yeah that happens though it happens because girls they jealous, they don't want you to do nothing, stuff like that."

[24] Tracking using friends is discussed here because it often overlapped with tracking using social media or other technology.

- Downstate, Senior High Boys

"I think it would be yeah because like, you're, you have a boyfriend and stuff, they would think that you were cheating on 'em so they would have one of their friends follow you to see what you were doing."
- Rural, Middle School Girls

"Say like in the lunch room, they def have a friend just looking at you to make sure you're not sitting with another guy or another girl or someone."
- Rural, Senior High Girls

"I know people like say one person like the girlfriend goes to one school and the boyfriend goes to another they ask people that they know at that school what has so and so been doing. And they like, having keep tabs on them."
- Upstate Urban 1, Senior High Girls

In addition to friends, cell phones can also be used to keep track of a dating partner's activities,

"A lot of Apple products have tracking devices in them…so you could easily find out where someone is and like, just show up."
- Rural, Senior High Boys

"R4: Like when you're texting, they want… they want…

R5: {interrupting}they wanna say 'why you're doing…'

{M: What you're doing. So they can use the texting…}

R1: {interrupting}where you are."
- Rural, Middle School Boys

In terms of using social media to follow a dating partner, there was robust discussion of how a partner's Facebook page or Twitter feed could be used to monitor activities; however, although both senior high girls and boys discussed that this was a common practice, some also felt there was a point at which it crossed the line, specific to frequency of use,

"When it's not ok is when every ten seconds I see that I'm logged in from another computer and I'm not logged in. That's when you cross a line, how are you constantly logged on my page but I'm not on your page, that's crossing the line for me."
- Upstate Urban 2, Senior High Boys

"I think its okay if like you know you checkin' if like how do I explain this…if like you have like a real purpose like oh, you wanted to see like what his friends is talking about chu or like you wanna like to find out like whether or not like…like...say like, oh my gosh like how did Jessica or like {laughs} we have a conversation and like they said like they cheated on you or

like {laughs} no but I mean like, its like kinda weird if you like, every second like in a day check to see what are they doing, who are they wit. Like it's just, it's just weird."
- *Upstate Urban 1, Senior High Girls*

Related to tracking on social media, password sharing was discussed as a common practice by senior high girls, senior high boys and middle school girls (the middle school boys in this sample said they didn't really share passwords):

"{M: Do you think it's common for people your age to give their Twitter or Facebook password?}

R1: Yea {multiple people agreeing with yea}

{everyone talking over each other about what happens if you do not give your password}

R: It's like a contract."
- *Upstate Urban 2, Senior High Boys*

"I understand if you're my boyfriend and I trust you but I don't know about …like, if I know you for mad long and we just started going out I'll give you my password because I trust you."
- *Downstate, Senior High Girls*

While sharing passwords appeared common, it was also recognized that once a dating partner had your password, they could use this to track or isolate you,

"I feel like they could, cuz like, they be like, they like make you give your password to your Twitter and stuff like that and moderate your conversations with people…"
- *Upstate Urban 2, Senior High Boys*

"R1: Or, if they like, if they have you password to your, like, Twitter and stuff, they could like, mess with it and like people unfollow you or you unfollow other people or like delete important things.

R3: Delete and block important people because…they don't want you to talk to them. Any communication, basically."
- *Upstate Urban 2, Senior High Girls*

"Oh it'd be easy because most relationships they like got your partner Facebook page stuff anyway so they just type in their password and see who she be messaging and who he be messaging in general or whatever."
- *Upstate Urban 1, Senior High Boys, in reference to how social media can be used to keep track of someone*

The large number of specific examples around how friends, cell phones and social media

could be used to monitor a dating partner's activities were incorporated into the 10C revision, and were also used to add a number of questions that focused on technology to other sections of the survey. Thus, a novel feature of the PATS is the inclusion of multiple questions on the use of technology in the perpetration of psychological aggression. Indeed, through the focus group conversations, participants emphasized the importance of technology in their lives, as evidenced in this conversation in a group of senior high girls,

"R2: That's really all for me. I think texting, computers, like social networking, that's basically what circulates our life.

R1: All of the teenagers, youth, that is a good percentage of what we do, social networking, texting, that's what your significant other is going to watch for.

R2: It's just the generation about technology.

R1: It always is.

{M: So, social networking, any other thoughts on technology?}

R3: I feel like…in our generation technology is more aggressive than actually being with each other; like, like you can...you can ruin somebody's life on technology than like for example cyber bullying, your partner can do that to you very easily.

…

R1: The bruises they can go away but the memory of how, of what happened …

R3: It haunts you.

R1: And then remembering all the things that people said about you…

{M: Right. And if it's on social media, you'll see it forever.}

R1: In seven years…it'll be here forever, so…

R3: They see everything, once you put something on the internet it stays forever and you delete it...it's always going to be there.

R1: It'll always come back to you somehow."
- *Rural, Senior High Girls*

Sex and contraceptives. Another large topic of discussion was on sex and

contraceptives, specifically in response to an investigator-added item ("kept you from getting the medical care you needed"), the *Wound Through Sexuality* group (7A – 7C) and question 14C ("tried to make a major decision that affected you without consulting you").

In regards to keeping a partner from getting medical care, the original question was viewed as too general, and not developmentally relevant. For example, one senior high girl (*Rural*) said "I put red because teenagers, they, they…I don't feel like they should have…shouldn't have any authority. I don't, you don't see a lot of that," and a senior high boy (*Upstate Urban 2*) said "I feel it's a very adult decision." However, participants in the first of the ten focus groups discussed how this question might be more relevant,

"R1: Yea, that's why I put blue cuz…like I mean, with like medical care, like, 'oh I'm dying' … 'don't go to the doctor,' but with things like 'oh I don't want you to take birth control,' 'oh, I want you to stop taking birth control,' or those aspects.

R2: {interrupts R1} I'll leave you if you go on birth control.

R3: Now, that is a completely different scenario."
- *Rural, Senior High Girls*

Because this seemed like a relevant adaptation, all subsequent groups were asked if the question would be improved by making it specific to contraceptive use. The majority of the senior high participants agreed that this would improve the question, and that preventing contraceptive use was a way you could hurt a dating partner; however, there was more agreement among the senior high girls about the relevance of this item than among the senior high boys.

"Yeah, especially if you're young. And you have, alright, if you feel like you need to be on birth control, you should because like both of y'all young and if you have a child, you're not gonna be able to take care of the child, then that's like next level…"
- Downstate, Senior High Girls, in response to whether a dating partner might keep you from getting birth control to harm you

"It's just not right, anyway. You shouldn't be kept from things you need. I mean, birth control isn't entirely necessary, but if you feel you should be on it, then you should be on it…it's not up to the other person."
- Rural, Senior High Boys, explaining why he thought someone his age wouldn't prevent a dating partner from using birth control

Middle school girls did not see this adaptation as an improvement, and disagreed that someone their age would prevent contraceptive use, citing their age (the discussion was not raised in the group of middle school boys due to time constraints).

> "I just don't think that would happen 'cuz we're so young."
> - *Rural, Middle School Girls*

Several of the senior high groups (*n*=2 girls groups, *n*=3 boys groups) also discussed birth control and abortion as part of question 14C (i.e., major decisions).

> "Um, I would look at a girl's perspective, if they was like, getting an abortion or something."
> - *Upstate Urban 2, Senior High Girls*

> "Like if she wanna use a condom or don't use a condom, you know, like that could be a major decision."
> - *Upstate Urban 2, Senior High Boys*

Finally, the questions in group 7 (especially 7B and 7C) led to rich discussion in all the senior high groups, and in the middle school girls group (these questions were not discussed in the middle school boys group due to time constraints; however, in the sticker data, the majority of middle school boys indicated that these questions were not applicable, with approximately 80% red stickers for both 7B and 7C). Two groups of senior high girls (rural and upstate urban) thought that teenage boys would not refuse sex (as described by 7B) – per one participant, "they're hormonal little teenage boys" (*Rural*). Another group thought perhaps the partner refused because "maybe they just [not] emotionally ready to have sex with you" (*Upstate Urban 1*), while a girl in another group asked "how do you know the relationship is even having sex?" (*Downstate*). Finally, one girl wondered "why would you even care" if someone refused to have sex with you (*Upstate Urban 1*). One female participant suggested that instead of refusing, teenage boys might call you names in order to make you feel insecure or inadequate,

> "I think that the only thing teenage boys would do is like, you're a whore….I'm not gonna sleep with you, you know. Something to make, oh, I'm gonna go sleep with your best

friend, she's not wide."
- *Rural, Senior High Girls*

Another girl from the same site suggested that "like, they kind of get more mad at you if you refuse to have sex with your partner," and this was indeed the consensus among the senior high boys, who suggested that "when you're a teenager, you're not gonna say no to sex, period" (*Upstate Urban 1*), but that a girl could "tease you, and come to the door with no clothes on, and that would be to piss you off" (*Upstate Urban 2*). Another participant suggested that a girl might refuse to have sex with you "just to make you feel insecure" (*Downstate*). One boy discussed that,

"For some teens, ya know, refusing to have sex is really you know some of, they probably have like, you know, a person they can really go to for sex, if like their partner's not really, you know, refusing, so if they say like 'if she, if he or she is refusing to have sex, I'll just go to this person for sex.'"
- *Upstate Urban 1, Senior High Boys*

Thus, while for boys a partner could refuse to make them feel insecure or inadequate, for girls, it appeared less about refusing, but instead about calling them names or insinuating that you could go elsewhere for sex. Because this question appeared to be viewed differently by males and females, several revisions were made to 7B (Appendix 5.A), focusing on the examples discussed by both males and females. Based on the observation that not all couples were having sex (and thus these questions may not apply), an N/A box was added to the response options.

In regards to 7C, there was some confusion over the word belittling, or how you could belittle someone in a sexual context, since the original question provided no specific examples. However, participants provided several helpful definitions, including,

"Like if they were mad at you, they could, like try to put you down, make them feel like they have power. Basically like bullying, you know?"
- *Rural, Middle School Girls*

"It can make the person feel bad about themselves, not good enough."
- *Downstate, Senior High Girls*

> "Something you're not comfortable with that you're doing, like, anything that you're doing in the bedroom, to humiliate you."
> - *Rural, Senior High Girls*

> "They're probably making fun of your penis."
> - *Upstate Urban 2, Senior High Boys*

> "Having sex in a demeaning way to the other person."
> - *Upstate Urban 1, Senior High Boys*

> "To make someone feel less of themselves in a way. So, you may not have sex and that may make your partner feel as if you were in a higher power than them in a way, not having sex with them."
> - *Downstate, Senior High Boys*

These definitions were used to create several revisions to 7C, such as "made you have sex in a way that you found demeaning, in order to humiliate you."

Summary. The above themes cover the major revisions that were made to the MPAB, in order to increase its relevance for teen dating relationships. Other revisions were made based on discussion about the remaining 26 questions, as well as discussion of questions that participants thought were missed but should be added. As with the previous 19 questions, these revisions primarily involved providing examples or making questions more specific. In order to ensure items were capturing purposeful and more egregious psychological aggression, the intent stems from the original MPAB were kept (in the original or a revised format), and intent was included in any new question.

Format

From these revisions, a preliminary list of 154 questions was generated for expert review and pilot testing (Appendices 5.A and 5.B). Table 5.3 lists the number of questions for testing per behavior type. The 154 questions include both those revised from the original MPAB, as well as adapted questions from other measures of psychological aggression that seemed to capture similar constructs (see Appendix 5.A). This large number of preliminary questions was created based on the suggestions of DeVellis (2003), who discusses that in the preliminary phases of

scale development, a large number of redundant items is desirable, in order to compare the relevance of different items that tap a similar construct, and choose the item that best captures the phenomenon. DeVellis (2003) also suggests that the scale developer include many more items in the preliminary scale than is desired in the final scale (in the current version of the PATS, there are approximately three times the number of final desired questions), as this allows the selection of the most internally consistent items. The Flesch-Kincaid reading level of the 154 items is 8.2 (similar to the original MPAB, which had a Flesch-Kincaid reading level of 8.0). This reading level may be a bit advanced, given that the majority of 8th and 12th grade students in the U.S. are able to read at only a basic level (National Center for Education Statistics, 2011), and a report looking at the top 40 books for students in grades 9-12 found that the average reading level of those books was 5.6 (Renaissance Learning, 2013). However, adolescents in the focus groups were able to comprehend the items, and where they had difficulty, revisions were made based on their suggestions; however, the reading level of the final scale should be kept in mind during expert review and pilot testing.

Table 5.3
Number of Revised Questions per Behavior Type

Behavior Type	Number of Preliminary Items
Sadistic	8
Threats	9
Isolating	16
Manipulation	10
Public humiliation	12
Verbal abuse	10
Wound through sexuality	15
Treat as inferior	9
Hostile environment	10
Monitoring	17
Wound through fidelity	7
Jealousy	10
Withheld emotional/physical	6
Control personal decisions	15
TOTAL	**154**

The preliminary PATS survey is divided into four sections: dating history screen, behavioral items, contextual prompts and overall assessment. Each section is described briefly below, and the full PATS survey is contained in Appendix 5.B. The PATS survey is intended to be administered as a computer-based instrument, in order to facilitate skip patterns and reduce participant burden.

Dating history screen. In order to help participants identify which dating relationship they should reference when answering the PATS, as well as to screen out participants who have never dated, a dating history screen was adapted from Wolfe et al. (2001). The screen assesses the type and number of dating relationships the participant has had (where dating is defined as *when you like a guy/girl, and he/she likes you back, and other people know you are together. This does not have to mean going on a formal date*; Giordano, Longmore, & Manning, 2006, p. 268), and guides the participant through a series of questions about the specific relationship they will think of when answering the behavioral items (a relationship that occurred in the past 12 months, whether current or recent).

Behavioral items. These are the 154 items described above. Following expert review and pilot testing (see *Future Directions*, below), this number should be reduced by about two-thirds. The response options for this section are *yes* (this has happened in my dating relationship), *no* (this has not happened in my dating relationship) and *not applicable*. The instructions to this section remind the participant to think specifically about the relationship he/she identified in the dating history screen, that their data is confidential, and that they can stop participating at any time.

Contextual prompts. Given the discussion in the Introduction about the importance of context in understanding the nature and impact of psychological aggression (which one focus group participant succinctly stated as "It matters how you say things. It depends on the

relationship" (*Downstate, Senior High Girls*), the PATS survey is designed to probe for additional information about selected behavioral items. If a participant selects *yes* to any behavioral item, a screen will pop up that asks them questions about that specific behavior, such as when the behavior occurred, where it occurred, who was present when it occurred, frequency with which it occurred, length of relationship prior to first incident, how the behavior made the participant feel and if the participant thought their partner was doing the behavior to be playful or as a joke. The computerized format allows these questions to be seen as soon as the participant endorses an item, and also allows for tailoring of questions. For example, the question asking about emotional impact of the behavior (Item G) can be made specific to the intent of the behavioral item – for the question "made you do sexual things you were not comfortable with, in order to embarrass or humiliate you," the emotional impact item would be written as "in general, how much did this experience embarrass or humiliate you?" The computer-based format is designed to reduce participant burden by only showing contextual items for endorsed behaviors; because of the severe nature of the items, it is expected that participants will not endorse a large number of items, although this assumption will be checked in quantitative testing. In order to further reduce burden, testing of the contextual items will be performed after the initial pilot study (see *Future Directions*, below), once the item pool is reduced. Contextual items were adapted from Mac Neil (2010) and Follingstad and Edmundson (2010).

Overall assessment. At the end of the behavioral items, all participants will be asked to indicate "overall, how psychologically abusive did/do you consider this partner," on a 4-point Likert-type scale (1=*not psychologically abusive at all*, 4=*very psychologically abusive*). This question was adapted from Follingstad and Edmundson (2010). After answering this question, a screen will pop up thanking the participant for his/her time, and providing information on local resources.

Discussion

The ability with which participants in this study were able to discuss psychological aggression in dating relationships supports our *a priori* belief that psychological aggression is in fact a common experience in adolescent romantic relationships, and one deserving of measurement. Their opinions on the original MPAB also support our belief that many of the adult items need revision before use with adolescent populations, including to encompass the importance of technology and electronic forms of aggression.

Although these participants were not recruited based on dating violence history, a number reported experience with physical and/or psychological aggression in dating relationships. This finding adds to literature highlighting the pervasiveness of teen dating violence in the lives of American adolescents, and also suggests that recommendations made by teens in this sample often came from personal experience. Many participants were also able to describe psychologically aggressive actions experienced by a friend, again suggesting that many of their thoughts and opinions were grounded in experience.

Social media use was essentially ubiquitous among these participants. The most commonly used social media sites were Facebook and Twitter, which aligns with a recent survey by investment bank and asset management firm Piper Jaffray, finding that of the 5,200 teens surveyed, Facebook was selected as the most important social network, followed closely by Twitter (Crook, 2013). However, when asked to describe the most important social media website, Facebook and YouTube were closely ranked as the most popular (with the popularity of Facebook declining since Spring 2012), followed by Twitter and Instagram. Additionally, teens in this survey noted that many of the social media sites they used weren't on the list of response options, highlighting the ever-changing nature of the social media landscape (Crook, 2013). Because of this, the majority of questions on the PATS were written without specifying a

specific social media name, so that the survey would remain relevant as popularity of these sites changes, and new sites emerge.

In addition to technology, topics about sex were also heavily discussed in the focus groups. Given that adolescent romantic relationships are important contexts for sexual development (Collins, 2003), this focus is not surprising. In Sullivan's (1953) work on adolescent psychosocial development, he specifically discusses that the romantic partner may influence the adolescent by his or her reaction during initial sexual experiences, with a negative reaction potentially damaging self-esteem, an idea referenced by many of the adolescents in our sample. Given the (understandable) importance of sex in the adolescent context, a number of questions were added to the PATS on this topic.

Finally, although we initially hypothesized that items on the revised MPAB would be relevant for middle school students (since approximately 25% of 12-year olds have had a dating relationship; since dating increases during the period from 12-17; and since perpetration of dating violence is reported by middle school students; Collins, 2003; Karriker-Jaffe, Foshee, Ennett, & Suchindran, 2008), we had difficulty recruiting middle school students, and the students that were recruited generally had more trouble discussing why a particular question was or was not relevant when compared to their senior high peers. For the middle school boys group, this may have been due to a language barrier (several of the participants had difficulty comprehending English) and time constraints, though these same limitations did not apply to the middle school girls group. Both were also from a rural site, and so it is possible that middle school youth from an urban setting may have a different perspective on these issues. However, based on the experiences in these groups, it seems that the PATS survey may be most appropriate for adolescents in grades 9-12, although this is a topic that could be explored in quantitative testing.

Future Directions

We will now begin the process of refining and testing the PATS survey. Reviewing the scale development steps recommended by DeVellis (2003), we have completed Steps 1 to 3 (determine the measurement topic; generate an item pool; determine the format for measurement), and now will move to Step 4 (expert review of initial item pool), followed by Steps 5 to 8 (related to piloting the measure, and performing initial validation work). Expert review is used to establish face validity, as well as to refine items (Kline, 2000); researchers and practitioners will be contacted who have experience working in teen dating violence, and asked to provide their feedback on the PATS survey. Following this review, the refined pool of items will be piloted with a sample of Canadian students. During pilot testing, we will specifically investigate the severity ratings of items within each category, in order to select mild, moderate and severe items. Following pilot revisions, the PATS survey will be administered to a larger sample of Canadian adolescents for initial reliability and validity testing. Per the recommendations of Follingstad (2009), assessments of physical and sexual dating aggression will also be included in these latter investigations, in order to better understand the nature of experienced psychological aggression.

This next phase of quantitative testing will provide preliminary information on the factor structure, reliability and validity of the PATS; however, given the nuanced nature of psychological aggression, it is only a first step in a detailed investigation of this construct in teen dating relationships. Other topics for future study include using confirmatory factor analysis to assess gender invariance of scale structure, and item response theory to assess differential item functioning among males and females (e.g., to determine if the use of the same scale(s) and/or scoring is appropriate for both sexes); determining if the PATS can also be used to understand perpetration (and relationships between perpetration and victimization); identifying factors

related to potential over-reporting of victimization (Follingstad & Rogers, 2013); investigating the most optimal recall period, as some investigations with teens suggest that a one year timeframe may not be optimal for accurate reporting (Hilton, Harris, & Rice, 1998; Jouriles, McDonald, Garrido, Rosenfield, & Brown, 2005); and using content from social networking posts and messages to understand everyday use of psychological aggression. Once we identify the final structure, another important topic for investigation will be whether shorter form(s) of the PATS survey can be found that still capture the construct of interest, as the length of the proposed survey is likely too long for inclusion in most quantitative research.

Limitations

The beginning of this article outlined a number of issues in the measurement of psychological aggression, and the PATS is a step to answer some of these critiques, primarily by including detailed measures of context, such as perceived intent, perception and frequency. However, the PATS survey is still a self-report measure, which, while desirable for many reasons, may prove a method incapable of assessing psychological aggression adequately (e.g., because of issues with over-reporting). Because of the importance of self-report measures to the field of psychology, though, it seems important to investigate if more sensitive uses of this method (e.g., with computer-based technologies that allow probing) can provide a picture of psychological aggression that better reflects the individual's lived reality. For example, a more sensitive application of computer-based self-report has been successfully used in the field of non-suicidal self-injury (Whitlock, Exner-Cortens, & Purington, 2013), another nuanced and complex behavior, suggesting similar methods may work here. In order to more thoroughly investigate the appropriateness of self-report items, planned investigations with the PATS survey include work with survivors of teen dating violence, to determine the sensitivity and specificity of the measure, as well as more detailed investigations of response options and administration

(e.g., differences in reporting if administered as a semi-structured interview scored by interviewers vs. as a self-report). Similar debates and careful work have shaped the measurement of stressful life events (e.g., Almeida, Wethington, & Kessler, 2002; Kanner, Coyne, Schaefer, & Lazarus, 1981; Rowlison & Felner, 1988; Wethington, Brown, & Kessler, 1995) and psychological child maltreatment (e.g., Cicchetti, 1991; Glaser, 2002; McGee & Wolfe, 1991), and knowledge from these fields will also be used to guide future investigation of the PATS survey.

Related to the methods for this study, we acknowledge that the group stickering activity may have been subject to social desirability bias. To minimize this, we gave clear instructions to participants to complete the activity based on their own thoughts and opinions, and not the thoughts and opinions of others in the group, and we also monitored participants during the stickering task, to ensure they were not discussing their choices with other participants; if a participant did start talking to others in the group, we reminded them that they would have the opportunity to share their thoughts during the upcoming discussion (though very few participants needed this reminder). To further address this limitation, we also collected warm-up questionnaire data. The data from the warm-up questionnaires were similar to the data from the stickering task, with six of the 19 revision questions identified on both tasks, and only three questions identified as needing revision on the warm-up questionnaire only. Rather, more questions were identified as needing revision on the sticker task alone (n=10), suggesting that if there was social desirability bias during the stickering activity, it resulted in participants being more likely to indicate that an action was *not* something someone their age would do. Because the purpose of this study was to gather comprehensive information on as many MPAB questions as possible, however, this bias likely did not affect the results in a negative way.

Conclusions

Psychological aggression is an important topic for researchers interested in adolescent romantic relationships, and given the importance of romantic partners in the adolescent context (Collins, 2003; Exner-Cortens, 2013), the potential for these aversive interpersonal experiences to lead to adverse outcomes should not be ignored; however, more work is needed to understand what qualifies as aversive and harmful in the teen context. This is especially important as adolescence is the period when individuals learn to negotiate romantic relationships, and so much of the currently assessed behavior may be considered normative and not harmful over the long-term. The PATS is an initial step in defining the construct of adolescent psychological aggression, by focusing on items described by adolescents as relevant, and specifically probing for context. Future testing will be important to refining and improving this preliminary measure.

References

Almeida, D. M., Wethington, E., & Kessler, R. C. (2002). The Daily Inventory of Stressful Events: An interview-based approach for measuring daily stressors. *Assessment, 9*, 41-55.

Baldry, A. C. (2003). "Sticks and stones hurt by bones but his glance and words hurt more": The impact of psychological abuse and physical violence by current and former partners on battered women in Italy. *International Journal of Forensic Mental Health, 2,* 47-57.

Basile, K. C., Arias, I., Desai, S., & Thompson, M. P. (2004). The differential association of intimate partner physical, sexual, psychological, and stalking violence and posttraumatic stress symptoms in a nationally representative sample of women. *Journal of Traumatic Stress, 17*, 413-421.

Black, M. C., Basile, K. C., Breiding, M. J., Smith, S. G., Walters, M. L., Merrick, M. T., …, & Stevens, M. R. (2011). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 summary report.* Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Cicchetti, D. (1991). Defining psychological maltreatment: Reflections and future directions. *Development and Psychopathology, 3*, 1-2.

Collins, W. A. (2003). More than myth: The developmental significance of romantic relationships during adolescence. *Journal of Research on Adolescence, 13,* 1-24.

Collins, W. A., Welsh, D. P., & Furman, W. (2009). Adolescent romantic relationships. *Annual Review of Psychology, 60*, 631-652.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., & Smith, P. H. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23*, 260-268.

Crook, J. (2013, April). *Facebook still reigns supreme with teens, but social media interest*

dwindling. Retrieved from http://techcrunch.com/2013/04/10/facebook-still-reigns-supreme-with-teens-but-social-media-interest-dwindling/

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd Ed.). Newbury Park, CA: SAGE Publications.

Draucker, C. B., & Martsolf, D. S. (2008). The role of electronic communication technology in adolescent dating violence. *Journal of Child and Adolescent Psychiatric Nursing, 23*, 133-142.

Emelianchik-Key, K. M. (2010). *The initial development and validation of the Teen Screen for Dating Violence*. (Unpublished doctoral dissertation). Old Dominion University, Norfolk, VA.

Exner-Cortens, D. (submitted manuscript). Theory and teen dating violence victimization: Considering adolescent development. *Developmental Review.*

Exner-Cortens, D., Eckenrode, J., & Rothman, E. (2013). Longitudinal associations between teen dating violence victimization and adverse health outcomes. *Pediatrics*, *131*, 71-78.

Feiring, C., & Furman, W.C. (2000). When love is just a four-letter word: Victimization and romantic relationships in adolescence. *Child Maltreatment, 5,* 293-298.

Furman, W., & Shaffer, L. (2003). The role of romantic relationships in adolescent development. In P. Florsheim (Ed.), *Adolescent romantic relations and sexual behaviors: Theory, research, and practical implications* (pp. 3-22). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Follingstad, D. R. (2007). Rethinking current approaches to psychological abuse: Conceptual and methodological issues. *Aggression and Violent Behavior, 12*, 439-458.

Follingstad, D. R. (2009). The impact of psychological aggression on women's mental health and behavior: The status of the field. *Trauma, Violence, & Abuse, 10*, 271-289.

Follingstad, D. R. (2011). A measure of severe psychological abuse normed on a nationally representative sample of adults. *Journal of Interpersonal Violence, 26*, 1194-1214.

Follingstad, D. R., Coyne, S., & Gambone, L. (2005). A representative measure of psychological

aggression and its severity. *Violence and Victims, 20*, 25-38.

Follingstad, D. R., & Edmundson, M. (2010). Is psychological abuse reciprocal in intimate relationships? Data from a national sample of American adults. *Journal of Family Violence, 25*, 495-508.

Follingstad, D. R., & Rogers, M. J. (2013). Validity concerns in the measurement of women's and men's report of intimate partner violence. *Sex Roles*, *69*, 149-167.

Follingstad, D. R., Rutledge, L. L., Berg, B. J., Hause, E. S., & Polek, D. S. (1990). The role of emotional abuse in physically abusive behaviors. *Journal of Family Violence, 5*, 107-120.

Giordano, P. C., Longmore, M. A., & Manning, W. D. (2006). Gender and the meanings of adolescent romantic relationships: A focus on boys. *American Sociological Review, 71*, 260-287.

Giordano, P. C., Manning, W. D., & Longmore, M. A. (2006). Adolescent romantic relationships: An emerging portrait of their nature and developmental significance. In A. C. Crouter & A. Booth (Eds.), *Romance and sex in adolescence and emerging adulthood: Risks and opportunities* (pp. 127-150). Mahwah, NJ: Lawrence Erlbaum Associates.

Glaser, D. (2002). Emotional abuse and neglect (psychological maltreatment): A conceptual framework. *Child Abuse & Neglect, 26*, 697-714.

Gondolf, E. W., Heckert, D. A., & Kimmel, C. M. (2002). Nonphysical abuse among batterer program participants. *Journal of Family Violence, 17*, 293-314.

Halpern, C. T., Oslak, S. G., Young, M. L., Martin, S. L., & Kupper, L. L. (2001). Partner violence among adolescents in opposite-sex romantic relationships: Findings from the National Longitudinal Study of Adolescent Health. *American Journal of Public Health, 91*, 1679-1685.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (1998). On the validity of self-reported rates of interpersonal violence. *Journal of Interpersonal Violence, 13*, 58-72.

Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative*

Health Research, 15, 1277-1288.

Jordan, C. E., Campbell, R., & Follingstad, D. (2010). Violence and women's mental health: The impact of physical, sexual, and psychological aggression. *Annual Review of Clinical Psychology, 6*, 607-628.

Jouriles, E. N., Garrido, E., Rosenfield, D., & McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress. *Child Abuse & Neglect, 33*, 451-460.

Jouriles, E. N., McDonald, R., Garrido, E., Rosenfield, D., & Brown, A. S. (2005). Assessing aggression in adolescent romantic relationships: Can we do it better? *Psychological Assessment, 17*, 469-475.

Kanner, A. D., Coyne, J. C., Schaefer, C., Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine, 4*, 1-39.

Karriker-Jaffe, K. J., Foshee, V. A., Ennett, S. T., & Suchindran, C. (2008). The development of aggression during adolescence. Sex differences in trajectories of physical and social aggression among youth in rural areas. *Journal of Abnormal Child Psychology, 36*, 1227-1236.

Katz, J., & Arias, I. (1999). Psychological abuse and depressive symptoms in dating women: Do different types of abuse have differential effects? *Journal of Family Violence, 14*, 281-295.

Kessler, R. C., Birnbaum, H., Bromet, E., Hwang, I., Sampson, N., & Shahly, V. (2010). Age differences in major depression: results from the National Comorbidity Survey Replication (NCS-R). *Psychological Medicine, 40*, 225-237.

Kessler, R. C., & Walters, E. E. (1998). Epidemiology of DSM-III-R major depression and minor depression among adolescents and young adults in the National Comorbidity Survey. *Depression and Anxiety, 7*, 3-14.

Kline, P. (2000). *A psychometrics primer*. London, UK: Free Association Books.

Krueger, R. A., & Casey, M. A. (2009). *Focus groups: A practical guide for applied research*. (4th Ed.). Thousand Oaks, CA: Sage Publications.

Lawrence, E., Yoon, J., Langer, A., & Ro, E. (2009). Is psychological aggression as detrimental as physical aggression? The independent effect of psychological aggression on depression and anxiety symptoms. *Violence and Victims, 24*, 20-35.

Leisen, M. B. (2000). *Development and validation of the Adolescent Partner Aggression Scale (APAS)*. (Unpublished doctoral dissertation). University of Hawai'i at Mānoa, Honolulu, HI.

Mac Neil, B. A. (2010). *Adolescents' experiences of psychologically aggressive behavior in their dating relationships and negative emotional and behavioral adjustment.* (Unpublished doctoral dissertation). University of New Brunswick, Saint John, New Brunswick.

Marshall, L. L. (1996). Psychological abuse of women: Six distinct clusters. *Journal of Family Violence, 11*, 379-409.

McGee, R. A., & Wolfe, D. A. (1991). Psychological maltreatment: Toward an operational definition. *Development and Psychopathology, 3*, 3-18.

Murphy, C. M., & Cascardi, M. (1993). Psychological aggression and abuse in marriage. In R. L. Hampton, T. P. Gullotta, G. R. Adams, E. H. Potter III, & R. P. Weissberg (Eds.), *Family violence: Prevention and treatment, Vol 1* (pp. 86-112). Newbury Park, CA: Sage Publications.

Murphy, C. M., & Hoover, S. A. (1999). Measuring emotional abuse in dating relationships as a multifactorial construct. *Violence and Victims, 14*, 39-53.

Murphy, K. A., Smith, D. I., & Xenos, S. (2012). TREAD: A promising change-target for partner abuse prevention with adolescents. *Journal of Family Violence, 27*, 345-356.

National Center for Education Statistics. (2012). *Digest of Education Statistics: 2011.* Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012001

Offenhauer, P., & Buchalter, A. (2011). *Teen dating violence: A literature review and annotated bibliography* (DOJ Document No. 235368). Washington, DC: Library of Congress.

O'Leary, K. D. (1999). Psychological abuse: A variable deserving critical attention in domestic violence. *Violence and Victims, 14,* 3-23.

Outlaw, M. (2009). No one type of intimate partner abuse: Exploring physical and non-physical abuse among intimate partners. *Journal of Family Violence, 24*, 263-272.

Renaissance Learning. (2013). *What kids are reading: The book-reading habits of students in American schools.* Retrieved from http://doc.renlearn.com/KMNet/R004101202GH426A.pdf

Robert Wood Johnson Foundation. (2011, January). *Youth Relationships Survey Instrument.* Retrieved from http://www.rwjf.org/en/research-publications/find-rwjf-research/2011/01/youth-relationships-survey-instrument.html

Rowlison, R. T., & Felner, R. D. (1988). Major life events, hassles, and adaptation in adolescence: Confounding in the conceptualization and measurement of life stress and adjustment revisited. *Journal of Personality and Social Psychology, 55*, 432-444.

Sackett, L. A., & Saunders, D. G. (1999). The impact of different forms of psychological abuse on battered women. *Violence and Victims, 14*, 105-117.

Sadie. (2010). *Psychological abuse and DV*. Retrieved from http://pandys.org/forums/index.php?showtopic=153991

Shepard, M. F., & Campbell, J. A. (1992). The Abusive Behavior Inventory: A measure of psychological and physical abuse. *Journal of Interpersonal Violence, 7*, 291-305.

Sugarman, D.B., & Hotaling, G.T. (1991). Dating violence: A review of contextual and risk factors. In B. Levy (Ed.), *Dating violence: Young women in danger* (pp. 100-118). Seattle, WA: Seal Press.

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. New York: W.W. Norton &

Company.

Thompson, M. P., Basile, K. C., Hertz, M. F., & Sitterle, D. (2006). *Measuring intimate partner violence victimization and perpetration: A compendium of assessment tools [Section C11]*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Wethington, E., Brown, G. W., & Kessler, R. C. (1995). Interview measurement of stressful life events. In S. Cohen, R.C. Kessler & L.U. Gordon (Eds.), *Measuring stress: A guide for health and social scientists* (pp. 59-79). New York: Oxford University Press.

Whitlock, J., Exner-Cortens, D., & Purington, A. (submitted manuscript). Assessment of non-suicidal self-injury: Development and initial validation of the Non-Suicidal Self-Injury Assessment Tool. *Psychological Assessment.*

Wolfe, D. A., & Feiring, C. (2000). Dating violence through the lens of adolescent romantic relationships. *Child Maltreatment, 5*, 360-363.

Wolfe, D. A., Scott, K., Reitzel-Jaffe, D., Wekerle, C., Grasley, C., & Straatman, A. (2001). Development and validation of the Conflict in Adolescent Dating Relationships Inventory. *Psychological Assessment, 13*, 277-293.

CHAPTER SIX

TEEN DATING VIOLENCE MEASUREMENT AND OUTCOMES: CONCLUSION

Teen Dating Violence Measurement and Outcomes: Conclusion

We began this volume with a discussion of developmental theory and how this theory can be used to understand violent and aggressive behaviors in adolescent romantic relationships. A consideration of developmental theory seems to address some of the current controversies facing the field, including why males and females might use equivalent amounts of some forms of aggression and why teen dating violence is mutual in many relationships, and also provides promising directions for future research. In reference to the debate around women using intimate partner violence in adult relationships, McHugh, Livingston and Ford (2011) argue:

> …against the conceptualization of intimate violence as a single truth or as a debate between polarized positions…accepting either the view that intimate violence is unilateral, that is, men beating women, or the conclusion that interpersonal violence is gender symmetric, that is, equally and reciprocally utilized by men and women, limits our conceptual framework and results in tunnel vision. Rather, we conceptualize interpersonal violence as a complex, multifaceted, and dynamic aspect of human interaction that occurs in multiple forms and patterns. The experience and meaning of violence is viewed as being connected to both the relationship and the larger context in which the violence occurs. (p. 323)

We agree with this sentiment, and assert that for adolescents, this "larger context" includes the developmental context. As discussed by Exner-Cortens (2013), understanding teen dating violence from a developmental perspective will not only contribute to improved understanding of this behavior, but will also contribute to an understanding of adolescent development more broadly. Indeed, Feiring and Furman (2000) describe that "investigators studying victimization and those studying romantic relationships have had little contact with each other" (p. 293), which is unfortunate, since "romance and victimization are not mutually exclusive experiences" (p.

294). We strongly agree with Feiring and Furman's (2000) statement that an integration of these two fields is necessary, and believe it would lead to positive developments in knowledge and understanding for both fields.

In Chapter Two, Exner-Cortens, Eckenrode, Bunge and Rothman demonstrated that dating violence experienced during early and mid-adolescence was indirectly associated with intimate partner violence experienced 12 years later, through the experience of intimate partner violence at 5.5 year follow-up. This finding adds to a growing body of literature documenting the contribution of teen dating violence to risk for re-victimization, as well as to literature showing that adolescent romantic relationships can impact the quality of romantic relationships in adulthood, and supports practice-based knowledge regarding the often chronic nature of partner violence victimization. A strength of this study is its use of propensity scores; while other empirical work will not always be able to include the large number of variables used in this study, researchers are encouraged to think carefully about potential confounds between teen dating violence and outcomes of interest (beyond socio-demographics), and include the most important of these in their surveys. While Exner-Cortens, Eckenrode, Bunge and Rothman addressed an important research question, they were not able to fully explore risk pathways from dating violence to future re-victimization. A more thorough exploration of these pathways is an important direction for future research.

The inability to explore these pathways was limited in part by the teen dating violence items included in their data (National Longitudinal Study of Adolescent Health, Add Health), which covered only mild physical and psychological victimization and did not include any information on how the participant perceived the aggression; these items were adapted from the revised Conflict Tactics Scale (CTS2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996), with data collected in 1996. As was shown in Exner-Cortens, Gill and Eckenrode (Chapters Three and

Four), this was before any psychometric testing of this measure in an adolescent sample, and the particular wording of Add Health teen dating violence items was not included in either of the adolescent CTS psychometric studies. However, Add Health is certainly not the only study that suffers from this limitation. In their comprehensive review, Foshee and Reyes (2011) estimate that over 200 articles on dating violence were published between 2002 and 2009, while Exner-Cortens, Gill and Eckenrode found only 38 that used a tested measure and presented sex-stratified prevalence and/or descriptives on an attitude measure during the time period 2002-2012. Thus, it appears the majority of dating violence articles are not using psychometrically tested scales to provide prevalence estimates, and reasons for this choice are an important discussion point for the field. Other discussion points (especially for researchers interested in developing new measures) include a more detailed and developmentally-focused exploration of the construct of dating violence, including if this construct can be adequately assessed using behavioral or attitude items alone, how multi-method approaches can be incorporated, and even the best definition of romantic relationships to use when studying dating violence (Furman & Hand, 2006); the definition of romantic relationships (or which relationship(s) the participant should think of when answering behavioral items) currently varies greatly by study and measure, and even studies using Add Health data don't consistently use the same definition (some include only romantic relationships, and others, including Exner-Cortens, Eckenrode, Schrader and Rothman in Chapter Two, include both romantic and sexual relationships). As described by McHugh et al. (2011), future measurement should "develop conceptualizations of intimate abuse as [it] occur[s] in a relationship, in time, and in a larger socio-historical context" (p. 329), and we agree, though for adolescents, would add to this definition a consideration of developmental context, as well.

Definitional issues are especially salient for psychological aggression (Follingstad &

Rogers, 2013), which, due to its subjective nature, is a difficult construct to assess. In a sample of school-going Spanish females (mean age=18.53 years), Rodríguez-Franco et al. (2007) found that although 77.2% reported ever experiencing at least one psychologically aggressive behavior, only 6.2% of the sample reported that they had ever felt abused. Against this, data from Jouriles, Garrido, Rosenfield and McDonald (2009) showed that associations between psychological dating violence and relationship anxiety, trauma symptoms and depression in a sample of high school students were similar whether they considered all endorsements, or only those endorsements that were described as unpleasant or not playful, though for the latter group, some associations were slightly larger. Together, these studies emphasize the need for more careful exploration of the nature of psychological aggression, as current measures may fail to capture important differences.

Considering this need specifically for psychological aggression that is intentional and more severe, Exner-Cortens, Eckenrode, Schrader and Rothman (Chapter Five) presented an adaptation of the Measure of Psychologically Abusive Behavior (MPAB; Follingstad, 2011). This adaptation was guided by the opinions of adolescents, and so reflects behaviors deemed harmful in the teen context, including a number of behaviors that describe how sex and technology can be used to harm a dating partner. This adapted measure also contains a number of contextual follow-up questions, to clarify how aggression was used and perceived. While this measure now requires psychometric testing, it is a promising step towards improved understanding of adolescent psychological aggression.

The field of teen dating violence is faced with a number of interesting and important questions. The papers presented in this volume aimed to address three of these questions, and the findings from each can be used to guide future study. Results about the state of teen dating violence measurement seem particularly important to consider, as answers to the field's

questions rest on the quality of measurement. We hope that researchers will find the developmental perspective presented at the beginning of this volume useful and will use developmental theory to guide future study of these attitudes and behaviors, which will serve to develop a more holistic understanding of adolescent romantic relationships, including normative and non-normative development of this emerging task. By working to understand adaptive and mal-adaptive romantic relationship behaviors, we can improve these first romantic experiences for all adolescents, experiences with implications that resonate throughout the life course.

References

Exner-Cortens, D. (submitted manuscript). Theory and teen dating violence victimization: Considering adolescent development. *Developmental Review.*

Feiring, C., & Furman, W. C. (2000). When love is just a four-letter word: Victimization and romantic relationships in adolescence. *Child Maltreatment, 5*, 293-298.

Follingstad, D. R. (2011). A measure of severe psychological abuse normed on a nationally representative sample of adults. *Journal of Interpersonal Violence, 26*, 1194-1214.

Follingstad, D. R., & Rogers, M. J. (2013). Validity concerns in the measurement of women's and men's report of intimate partner violence. *Sex Roles, 69*, 149-167.

Foshee, V. A., & Reyes, H. L. M. (2011). Dating abuse: Prevalence, consequences, and predictors. In R. J. R. Levesque (Ed.), *Encyclopedia of Adolescence* (pp. 602-615). New York, NY: Springer Science + Business Media.

Furman, W., & Hand, L. S. (2006). The slippery nature of romantic relationships: Issues in definition and differentiation. In A. C. Crouter & A. Booth (Eds.), *Romance and sex in adolescence and emerging adulthood: Risks and opportunities* (pp. 171-178). Mahwah, NJ: Lawrence Erlbaum Associates.

Jouriles, E. N., Garrido, E., Rosenfield, D., & McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress. *Child Abuse & Neglect, 2009,* 451-460.

McHugh, M. C., Livingston, N. A., & Ford, A. (2005). A postmodern approach to women's use of violence: Developing multiple and complex conceptualizations. *Psychology of Women Quarterly, 29*, 323-336.

Rodríguez-Franco, L., Borrego, J. L., Rodríguez-Diaz, F. J., Molleda, C. B., Pineda, C. E., Bellerín, M. A. A., & Quevedo-Blasco, R. (2012). Labeling dating abuse: Undetected abuse among

Spanish adolescents and young adults. *International Journal of Clinical and Health Psychology, 12*, 55-67.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283-316.

APPENDIX

Appendix 2.A

Variables for Propensity Score Model (n=3,961)

Variable	Wave	M (SE) or % (n)	Description
Socio-demographic			
Age (Foster et al., 2004; Lichter & McCloskey, 2004)[c]	1	M(SE) = 15.47 (0.10), range: 11.39-18.53	Age at the time of the Wave 1 interview.
Sex (Spriggs et al., 2009a; Tschann, Pasch, Flores, Marin, Baisch, & Wibbelsman, 2009)[d]	1	• Male[a]: 46.3 (1617) • Female: 53.7 (2344)	Interviewer-confirmed sex at Wave 1.
Race/ethnicity (Halpern et al., 2009; Spriggs et al., 2009a)[e]	1	• White, non-Hispanic[a]: 70.9 (2318) • Black, non-Hispanic: 12.5 (714) • Hispanic: 10.1 (569) • Other: 6.5 (360)	Racial and ethnic background (self-reported).
Family structure (Foster et al., 2004; Halpern et al., 2009; Spriggs et al., 2009a)[e]	1	• Two biological parents[a]: 58.2 (2257) • Step-family: 17.2 (697) • Single parent: 21.8 (896) • Other (e.g., grandparents): 2.8 (111)	Roster of all people who lived in the participant's household (self-reported).
Parental education (SES) (Foster et al., 2004; Halpern et al., 2009; Spriggs et al., 2009a)[c]	1	M(SE) = 3.97 (0.06), range: 1-6	Highest education level for the resident father or mother. Measured on a 6-point Likert-type scale (1=*8th grade or less*, 6=*post-graduate training*).
Religiosity (Halpern, Oslak, Young, Martin, & Kupper, 2001[b]; Howard, Qiu, & Boekeloo, 2003[b])[c]	1	M(SE) = 3.03 (0.04), range: 1-4	Participants were asked how important religion was to them. Measured on a 4-point Likert-type scale (1=*not important at all*, 4=*very important*). Individuals who reported no religion were coded as a 1.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Socio-demographic			
Pubertal status (Foster et al., 2004)[f]	1	M(SE) = 0.058 (0.029), range: -2.17-2.03	At Wave 1, participants rated themselves on three indicators of physical maturity (e.g., facial hair growth for boys, degree of breast development for girls), similar to items found in the Pubertal Development Scale (Petersen, Crockett, Richards, & Boxer, 1988). Following Foster et al. (2004), each item was first standardized to mean 0 and SD 1 and then averaged to create the pubertal status score.
GPA (Cleveland, Herrera, & Stuewig, 2003; Halpern et al., 2001[b])[c]	1	M(SE) = 2.85 (0.03), range: 1-4; α=0.75	For participants who were in school, their mean grade point average (GPA) across four subjects (English, Math, Social Studies, Science). Measured on a 4-point scale (1=*D or lower,* 2=*C*, 3=*B,* 4=*A*).
Picture Vocabulary Test (Cleveland et al., 2003)	1	M(SE) = 103.20 (0.59), range: 36.00-138.00	Participant's standardized score on the Add Health Picture Vocabulary Test.
Individual			
Low self-esteem (Foshee, Benefield, Ennett, Bauman, & Suchindran, 2004; O'Keefe & Treister, 1998[b])	1	• Yes: 33.6 (1343) • No[a]: 66.4 (2616)	Self-esteem was assessed using four items from Rosenberg's self-esteem scale (e.g., "you like yourself just the way you are"; Rosenberg, 1965). Items were reverse coded and summed, so that higher scores indicate higher self-esteem (range, 0-16; α=0.78). Because of the skewed distribution of total self-esteem scores, scores were re-coded, so that 1=*participants in the lowest quartile* (total score ≤ 11).

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Individual			
Depression (Cleveland et al., 2003; Tschann et al., 2009)[c]	1	• Yes: 14.3 (603) • No[a]: 85.7 (3358)	Nine items from the 20-item Centers for Epidemiologic Studies—Depression Scale were used to assess depressive symptomatology (Radloff, 1977), asking if participants had experienced particular feelings in the past seven days (e.g., "you felt depressed"). The nine items were summed; higher scores indicate greater depressive symptomatology (range, 0-27; α=0.80). Using the suggestions of Primack, Swanier, Georgiopoulos, Land and Fine (2009), depression cut scores were created for males and females (using a cut-off of ≥ 11 for females and ≥ 10 for males). Cut scores were used in the propensity score analysis.
Anhedonia (Cleveland et al., 2003; Tschann et al., 2009)	1	• Yes: 11.1 (498) • No[a]: 88.9 (3463)	Sum of two items asking participants if they were happy or enjoyed life in the past seven days (Testa & Steinberg, 2010). Items were reverse scored so that higher scores indicate greater anhedonia (range, 2-8; α=0.70). Because of the skewed distribution of total anhedonia scores, scores were re-coded, so that 1=*participants in the highest quartile* (total score ≥ 6).
Self-efficacy[c]	1	M(SE) = 15.86 (0.05), range: 7-20; α=0.56	Sum of four items assessing individual self-efficacy (e.g., "you have a lot of energy"; Hitlin & Elder, 2007). Responses were assessed on a 5-point Likert-type scale (1=*strongly agree*, 5=*strongly disagree*). Items were reverse scored so that higher scores indicate greater self-efficacy.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Individual			
Planfulness	1	M(SE) = 15.11 (0.07), range: 4-20; α=0.74	Sum of four items assessing planfulness (e.g., "when you have a problem to solve, one of the first things you do is get as many facts as possible"; Hitlin & Elder, 2007). Responses were assessed on a 5-point Likert-type scale (1=*strongly agree*, 5=*strongly disagree*). Items were reverse scored so that higher scores indicate greater planfulness.
Optimism	1	• Low: 31.3 (1255) • Medium[a]: 31.9 (1250) • High: 36.8 (1454)	Participants indicated whether they thought they would go to college (1=*low likelihood*, 5=*high likelihood*), whether they thought they would live to age 35 (1=*almost no chance*, 5=*almost certain*), and whether they felt hopeful about the future (0=*never or rarely*, 3=*most or all of the time*; Hitlin & Elder, 2007). Individual items were standardized and summed. Because of the skewed distribution of total optimism scores, scores were re-coded by tertile, to indicate low, medium and high optimism scores.
Out-of-school time participation	1	• Both sport and non-sport: 42.6 (1240) • Non-sport only: 22.2 (712) • Sport only: 22.4 (633) • None[a]: 12.8 (415)	On the in-school survey, participants were asked if they participated in one of 12 after-school sports (e.g., basketball), one of 21 after-school non-sport activities (e.g., science club), or no after-school activities. Participation was divided into four categories based on Jiang and Peterson (2012).
Suicidal thoughts and behavior (Kim-Godwin, Clements, McCuiston, & Fox, 2009[b])[c,f]	1	• Thoughts only: 9.5 (394) • Thoughts and attempt(s): 4.5 (168) • None[a]: 86.0 (3399)	Categorical variable indicating if the participant seriously thought of committing suicide in the past year, and if so, if they also made at least one suicide attempt.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Individual			
Unhealthy weight control (Kim-Godwin et al., 2009[b])	1	• Yes: 0.7 (41) • No[a]: 99.3 (3920)	Dichotomous variable indicating if the participant vomited, took diet pills and/or took laxatives in the past seven days in order to lose weight or keep from gaining weight.
Heavy episodic drinking (Cleveland et al., 2003)[c,f]	1	• Yes: 9.2 (355) • No[a]: 90.8 (3606)	To assess drinking behavior, participants reported how many times they drank five or more drinks in a row in the past year. A dichotomous variable indicates whether participants had done this at least two to three days per month for the past year.
Smoking (Kim-Godwin et al., 2009[b])[c,f]	1	• Yes: 27.5 (1030) • No[a]: 72.5 (2931)	Dichotomous variable indicating any cigarette smoking (i.e., smoking on one or more days) in the past 30 days.
Marijuana use (Raiford, Wingood, & DiClemente, 2007)[c,f]	1	• Yes: 12.8 (511) • No[a]: 87.2 (3450)	Dichotomous variable indicating any marijuana use in the past 30 days.
Other drug use (Raiford et al., 2007)[c,f]	1	• Yes: 4.6 (171) • No[a]: 95.4 (3790)	Dichotomous variable indicating any cocaine, injection drug, inhalant drug and/or other illegal drug (e.g., LSD, PSP, ecstasy) use in the past 30 days.
Age of alcohol initiation (Kim-Godwin et al., 2009[b])[f]	1	M(SE) = 13.13 (0.09), range: 1-17	Age of participant when they had their first drink of beer, wine or liquor and weren't with their parents or other adults in their family. Variable excludes participants who reported they had never drank.
Physical fighting (Foshee et al., 2004)[c,f]	1	• Yes: 32.5 (1207) • No[a]: 67.5 (2754)	Dichotomous variable indicating whether the participant had been in a serious physical fight in the past year.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Individual			
Violent behavior (Cleveland et al., 2003)[f]	1	• Yes: 35.1 (1348) • No[a]: 64.9 (2612)	Sum of four items, indicating whether the participant had been in a physical fight, had been jumped, had pulled a knife or gun on someone, or had shot or stabbed someone in the past year (Hitlin & Elder, 2007). Because of the skewed distribution of total violent behavior scores (range, 0-8), scores were re-coded, so that 1=*participants in the highest quartile* (total score ≥ 1).
Delinquent behavior (Foster et al., 2004; Maas, Fleming, Herrenkohl, & Catalano, 2010)[c,f]	1	• No delinquency[a]: 55.7 (2253) • One delinquent act: 18.8 (759) • Two or more delinquent acts: 25.5 (949)	Seven items from the Self-Reported Delinquency scale assessed the frequency of delinquent behavior over the past year (e.g., deliberately damaging property that didn't belong to them, selling marijuana or other drugs; Elliot, Ageton, & Huizinga, 1985). Based on the strong positive skew of the sum score, variable was collapsed into no delinquency (*=0*), one delinquent act (*=1*), or two or more delinquent acts (*=2*).
Interpersonal			
Child maltreatment (Foshee et al., 2004; Wekerle & Wolfe, 1998)[e,f]	4	• Yes: 20.0 (806) • No[a]: 80.0 (3155)	Retrospective report of physical and/or sexual child maltreatment that occurred before the participant's 18th birthday (e.g., "how often did a parent or other adult caregiver touch you in a sexual way, force you to touch him or her in a sexual way, or force you to have sexual relations?"). Questions were similar to those in the Parent-Child Conflict Tactics Scale (Straus, Hamby, Finkelhor, Moore, & Runyan, 1998). A dichotomous variable indicates whether participants reported any form of abuse.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Interpersonal			
Forced sex (Foshee et al., 2004)[f]	1	• Yes: 6.0 (136) • No[a]: 94.0 (2203)	Female participants only were asked if they were physically forced to have sexual intercourse against their will by any person. A dichotomous variable reflects endorsement of forced sex by female participants.
Peer deviance (Howard et al., 2003[b]; Vézina et al., 2011[b])[c,f]	1	• Low: 36.6 (1486) • Medium[a]: 29.2 (1128) • High: 34.2 (1347)	Participants indicated how many of their three best friends smoked, drank and/or used marijuana, or if they had taken part in a fight where one group of friends was against another group in the past year. Based on Lee (2011), individual scores were standardized and then summed. Because of the skewed distribution of total peer deviance scores, scores were re-coded into three categories, based on tertiles.
Best friend influence	1	• Good[a]: 68.5 (2305) • Neither bad nor good: 27.7 (969) • Bad: 3.8 (117)	During the parent interview, the participant's parent indicated whether they thought their child's best friend was a good influence, a bad influence, or neither a good nor bad influence.
Social acceptance	1	• No: 13.1 (537) • Yes[a]: 86.9 (3423)	Participants were asked if they felt socially accepted (1=*strongly agree,* 5=*strongly disagree*). Because of the skewed distribution of total social acceptance scores, scores were re-coded, so that 1=*participants in the highest quartile* (total score ≥ 3).
Loved and wanted	1	• No: 8.9 (379) • Yes[a]: 91.1 (3582)	Participants were asked if they felt loved and wanted (1=*strongly agree,* 5=*strongly disagree*). Because of the skewed distribution of total loved and wanted scores, scores were re-coded, so that 1=*participants in the highest quartile* (total score ≥ 3).

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Interpersonal			
Family disadvantage (Spriggs, Halpern, Herring, & Schoenbach, 2009b)	1	• No disadvantage[a]: 47.9 (1589) • One disadvantage indicator: 34.1 (1252) • Two or more disadvantage indicators: 18.0 (643)	Sum of dichotomized family structure (0=*two biological parents*), receipt of public assistance (parent report), reported difficulty paying bills (parent report), parental education (1=*less than high school*) and parental unemployment (parent report). Based on the skewed distribution of the sum score, variable was collapsed into three categories.
Close relationship with mother (Cleveland et al., 2003)	1	• No: 32.1 (1283) • Yes[a]: 67.9 (2536)	Sum of four items indicating the quality of the participant's relationship with their mother (e.g., "most of the time, your mother is warm and loving toward you"), assessed on a 5-point Likert-type scale (1=*strongly agree,* 5=*strongly disagree*). Scores were reverse coded so that a higher score indicates a better relationship. Because of the skewed distribution of the total score (range, 4-20), scores were re-coded, so that 1=*participants in the lowest quartile* (total score ≤ 16).
Close bonding with parents (Maas et al., 2010)	1	• No: 23.4 (727) • Yes[a]: 76.6 (2068)	Mean of 10 items that indicated the quality of the participant's relationship with their mother and father (e.g., "overall, you are satisfied with your relationship with your [mother/father]"), assessed on a 5-point Likert-type scale (1=*strongly agree*, 5=*strongly disagree*). Scores were reverse coded so that a higher score indicates a better relationship. Because of the skewed distribution of the total score (range, 1-5), scores were re-coded, so that 1=*participants in the lowest quartile* (total score ≤ 4.4).

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Interpersonal			
Monitoring (Howard et al., 2003[b])[c]	1	M(SE) = 5.02 (0.05), range: 0-7; α=0.58	Seven-item scale, described by Jiang and Peterson (2012). Using a dichotomous response option (*yes*/*no*), participants reported on the level of parental monitoring in adolescent decision making (e.g., "do your parents let you make your own decisions about what time you go to bed on week nights?"). Higher scores indicate less parental monitoring.
Number of meals together, past week	1	• Three or less: 27.3 (1200) • Four or more[a]: 72.7 (2755)	Participants reported how many days in the past week at least one of their parents was in the room while they ate their evening meal. Item was recoded so that 1=*participants in the lowest quartile* (number days ≤ 3).
Social support[c]	1	M(SE) = 28.27 (0.10), range: 11-35; α=0.77	Seven-item scale, described in Wight, Botticello and Aneshensel (2006). Participants reported if they felt cared about by adults, teachers, parents and friends, and if they felt understood by, had fun with, and paid attention to by their family (1=*not at all*, 5=*very much*). Items were summed, so that higher scores indicate more support.
Ever sexually active (Kaestle & Halpern, 2005)[f]	1	• Yes: 32.5 (1366) • No[a]: 67.5 (2591)	Dichotomous variable indicating whether the participant had ever had penile-vaginal sexual intercourse.
Number of non-romantic sexual partners (Cleveland et al., 2003)[c,f]	1	• Three or more: 4.4 (180) • One or two: 10.6 (435) • None[a]: 85.0 (3346)	Categorical variable indicating the number of non-romantic sexual partners the participant had since January 1, 1994 (~1.5 years). Participants in the *no* category had either had no non-romantic sexual partners since January 1, 1994, or had never had sex.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
Interpersonal			
Early sexual debut (Halpern et al., 2009)[c,f]	3	• First sex age 15 or younger: 33.6 (1264) • First sex age 16 or older[a]: 66.4 (2697)	Dichotomous variable indicating whether the participant's first sex was before or after age 15. Age break was based on the lowest tertile (Halpern et al., 2009).
Number of romantic dating relationships (Halpern et al., 2001[b]; Halpern et al., 2009; O'Keefe & Treister, 1998[b])[c]	1	M(SE) = 1.15 (0.03), range: 0-3	Categorical variable indicating the number of romantic relationships the participant had in the prior 18 months. Participants were allowed to list up to three such relationships.
School/Community			
School attachment (Cleveland et al., 2003)[c]	1	M(SE) = 11.43 (0.07), range: 3-15; α=0.77	Sum of three items indicating how close the participant felt to his/her school (e.g., "you feel like you are a part of your school"). Score is for participants who were in school (i.e., excludes participants who were not in school). Responses were assessed on a 5-point Likert-type scale (1=*strongly agree*, *5=strongly disagree*). Items were reverse coded, so that higher scores indicate higher attachment.
School disadvantage (Spriggs et al., 2009b)[c]	1	M(SE) = 0.82 (0.03), range: 0.15-1.91	Aggregate of total family disadvantage (aggregated by school ID).
Older school mean age (Spriggs et al., 2009b)	1	• Yes: 13.2 (745) • No[a]: 86.8 (3216)	Aggregate of Wave 1 participant age (aggregated by school ID). Variable was re-coded so that 1=*participant in the highest school quartile* (mean age > 16.7).
School size (Halpern et al., 2001[b])	1	• Small[a]: 17.2 (665) • Medium: 47.1 (1554) • Large: 35.6 (1742)	School administrators were asked to indicate whether the school the participant attended was small (1-400 students), medium (401-1000 students) or large (1001-4000 students) in size.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
School/Community			
Parent involvement in school	1	• High: 20.1 (667) • Low[a]: 79.9 (2859)	School administrators were asked to indicate the approximate percentage of children in the school with a family involved in a parent organization. This variable was re-coded so that 1=*participant in highest school quartile* (percentage of parents involved > 40).
School below grade level	1	• Yes: 23.7 (734) • No[a]: 76.3 (2298)	School administrators were asked to indicate the approximate percentage of students in their school who were testing one or more grade levels below (based on standardized achievement tests). This variable was re-coded so that 1=*participant in highest school quartile* (percentage of students testing below grade level > 26).
Trouble at school	1	M(SE) = 0.02 (0.03), range: -1.41-2.09	On the in-school questionnaire, participants were asked four items, indicating trouble at school (e.g., "you have trouble getting along with teachers"; Hitlin & Elder, 2007). Responses were measured on a 5-point Likert-type scale (0=*never*, 4=*everyday*). The final summed scale was standardized, so that higher scores indicate more trouble relative to the participant's peers (α=0.85).
Witness violent crime (Spriggs et al. 2009a)[c,f]	1	• Yes: 10.0 (437) • No[a]: 90.0 (3524)	Dichotomous variable indicating whether the participant had seen someone shoot or stab another person in the past year.
Safe in neighborhood (O'Keefe & Treister, 1998)[c]	1	• No: 7.7 (367) • Yes[a]: 92.3 (3594)	Dichotomous variable indicating whether the participant usually felt safe in their neighborhood.

(Appendix 2.A continues)

(Appendix 2.A continued)

Variable	Wave	M (SE) or % (n)	Description
School/Community			
Low neighborhood social control (O'Keefe & Treister, 1998)	1	• Yes: 25.8 (962) • No[a]: 74.2 (2598)	Sum score of two items assessing neighborhood social control (e.g., "if your neighbor saw your child getting into trouble, would your neighbor tell you about it?"). Data were provided by parent report, and measured on a 5-point Likert-type scale (1=*definitely would*, 5=*definitely would not*). Items were reverse-coded so that higher scores indicate greater social control. Because of the skew in the total scores, the scale was re-coded so that 1=*participant in lowest quartile* (score ≤ 7).
High neighborhood disadvantage (O'Keefe & Treister, 1998)	1	• Yes: 41.6 (1473) • No[a]: 58.4 (2042)	Sum score of three items assessing neighborhood disadvantage (e.g., "in this neighborhood, how a big a problem is litter or trash on the streets and sidewalks?"). Data were provided by parent report, and measured on a 3-point Likert-type scale (*1=no problem at all*, 3=*a big problem*). Items were reverse-coded so that higher scores indicate less disadvantage. Because of the skew in the total scores, the scale was re-coded so that 1=*participant in lowest quartile* (score ≤ 7).

Percentages and means are weighted, number of subjects, range and Cronbach's alpha are unweighted.
Note: High school drop-out was also assessed as a school/community level variable, but in the final sample of 3,961, there were no positive endorsements of this variable, and so it was not included in any analyses.
[a]Reference group
[b]Cross-sectional study
[c]Included in propensity score model
[d]Exact match variable
[e]Mahalanobis match variable
[f]Administered via A-CASI

Appendix 3.A
Measurement Quality Assessment: Behavior Scales

Authors and year	Content	Sampling	Reliability	Validity
CTS				
Cascardi et al. (1999)	Psychological and physical violence victimization and perpetration	Students in mandatory health education classes (non-representative)	n/a	Construct (convergent)
Nocentini et al. (2011)	Physical violence perpetration	High school students (non-representative)	Internal consistency	Construct (factor structure, measurement invariance)
CADRI				
Fernández-Fuertes et al. (2006)	Physical, psychological, and sexual violence victimization and perpetration	Public high school students (non-representative)	Internal consistency	n/a
Fernández-Gonzalez et al. (2012)	Physical, psychological, and sexual violence victimization and perpetration [only perpetration scale tested]	Study 1: High school students (non-representative) Study 2: Child Protective Services-involved youth (representative)	Internal consistency	Predictive Concurrent Construct (factor structure, convergent)
Hokoda et al. (2006)	Physical, psychological, and sexual violence victimization and perpetration	High school students (non-representative)	Internal consistency Test-retest	n/a
Jouriles et al. (2005) and Jouriles et al. (2009)	Physical and psychological (threatening behavior, verbal/emotional) victimization	High school students (non-representative)	Internal consistency	Construct (convergent)

(Appendix 3.A continues)

(Appendix 3.A continued)

Authors and year	Content	Sampling	Reliability	Validity
CADRI				
Wolfe et al. (2001)	Physical, psychological, and sexual violence victimization and perpetration [only perpetration scale tested]	Study 1: High school classrooms (non-representative) Study 2: Representative sample of 10 secondary schools (urban, semi-rural and rural) Study 3: One public secondary school (non-representative) Study 4: Local youth agency (non-representative)	Internal consistency (St 2) Test-retest (St 3) Partner agreement (St 3)	Construct (factor structure, St 2; convergent, St 4)
Other – Behavior				
Lavoie & Vézina (2001)	Physical, psychological, and sexual violence victimization (females) and perpetration (males)	Two low-to-middle SES public high schools (non-representative)	Internal consistency	Criterion Construct (discriminant)[a]
Leisen (1999)	Physical, psychological, and sexual violence victimization and perpetration	Study 1: Social service agencies and schools Study 2: Six schools and programs in four cities Study 3: Public high school (all non-representative)	Internal consistency (St 2/St 3)	Content (St 1) Construct (convergent, discriminant; St 3)

(Appendix 3.A continues)

(Appendix 3.A continued)

Authors and year	Content	Sampling	Reliability	Validity
Other – Behavior				
Murphy et al. (2012)	Warning sign behaviors (potentially harmful behaviors that may be indicative of more harmful abuse) and likely responses to those behaviors	Females in 10 low-to-middle SES secondary schools (non-representative)	Internal consistency Inter-rater	Content[b] Criterion Construct (factor structure)
Schultz & Jaycox (2008)	Fear experienced in dating relationships	9th grade health classes serving primarily Latino populations (representative)	Internal consistency Test-retest	Construct

Abbreviations: St – Study

[a]Discriminant validity data were collected on a sub-sample (*n*=144, 60.4% female). Criterion validity data were collected six months after the initial data collection, using interviews with 20 female participants and 28 male participants.

[b]Murphy et al. (2012) also present some results from a convenience sample of 426 young adult women (aged 18-25), recruited online. Data from this sample were used to assess content validity.

Appendix 3.B

Male Victimization Prevalence Estimates

Community-based samples									
							Victimization Prevalence (%)		
Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CADRI									
Archer et al. (2010)[3]	CADRI – Spanish	6 (P)[b]	601	58.2[a]	Spanish	In a relationship	25.2		
Fernández-Fuertes & Fuertes (2010)[3]	CADRI – Spanish	14 (4 P, 10 Ps)	567	58.4[a]	Spanish	In a relationship (past year)	26.3	94.5	
Schiff & Zeira (2005)[3]	CADRI – Hebrew	25 (4 P, 17 Ps, 4 S)	105	51.0	Israeli	Past year	18.0-30.6[c]	12.2-78.0[c]	12.2-34.0[c]
Wolfe et al. (2003)[2]	CADRI	n/a (P, Ps)	158[d]	50.0[a]	Primarily White	Past 2 months (current or recent dating partner)	33.0	29.0-43.0[e]	
CTS									
Black et al. (2008)[1]	CTS2 – modified	14 (P, S)	197	56.0[f]	Multi-racial (47% Black, 41% White)[f]	n/a	12.7[a,g]		
Boafo (2011)[3]	CTS2	5 (3 P, 1 Ps, 1 S)	3747	48.2	South African	Ever	9.4-16.9	11.0	15.0

(Appendix 3.B continues)

(Appendix 3.B continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Callahan et al. (2003)[1]	CTS2	14 (12 P – modified, 2 S)	190	47.4[a]	Multi-racial (50.5% Black, 49.5% White)[a]	Ever	36.0[h]		
Jiménez et al. (2008)[3]	CTS	9 (P)	446[i]	52.5	Spanish	Within the couple (past 6 months)	37.9[j]		
Lichter & McCloskey (2004)[1]	CTS	5 (P)	208	51.0	n/a[k]	Past year	1.9-11.2[l]		
Muñoz-Rivas et al. (2007)[3]	M-CTS – Spanish	13 (8 P, 5 Ps)[m]	2416[n]	58.6	Spanish	Actual partner	31.3	92.3	
O'Leary & Slep (2003)[1]	M-CTS – Adolescent	10 (P)[o]	206[p]	58.3[a]	Primarily White (63.6%; 15.5% Black, 8.3% Hispanic)[a]	Duration of current dating relationship	24.4		

(Appendix 3.B continues)

(Appendix 3.B continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
O'Leary et al. (2008)[1]	M-CTS – Adolescent	14 (8 P, 6 Ps)	2363[p]	49.8[a]	Multi-racial (55.1% White, 14.0% Black, 17.9% Hispanic)	Duration of current / most recent relationship	31.0	85.0	
Sears et al. (2007) and Sears & Byers (2010)[2]	CTS – modified	7 (3 P, 4 Ps)	633[q]	48.8[a]	Primarily White	Ever	41.0	59.0	
Probability/census samples									
CADRI									
Pradubmook-Sherer (2009)[3]	CADRI – Thai	23 (4 P, 15 Ps, 4 S)	1296[r]	54.5	Thai	n/a	41.9[s]	49.2-65.8[s]	43.2[s]
Sherer (2009) and Sherer & Sherer (2008)[3]	CADRI – Hebrew & Arabic	23 (4 P, 15 Ps, 4 S)	1357[t]	56.4[a]	Israeli (47.2% Jewish, 52.8% Arab)	Current/most recent partner	41.4	35.3-88.9[u]	46.4

(Appendix 3.B continues)

(Appendix 3.B continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Giordano et al. (2010) and Richards & Branch (2012)[1]	Revised CTS	4 (P)	970[v]	51.0[a]	Primarily White (64.0%; 23.0% Black, 11.0% Hispanic)	Current/most recent dating partner	29.3[a]		
Swart et al. (2002)[3]	CTS2	20 (P)	928	53.2[a,w]	South African	Past year	37.8		
Other									
Ulloa et al. (2004)[1]	RFDE	11	678[x]	55.0	100% Hispanic	Past 6 months		67.0	

Abbreviations: P – Physical; Ps – Psychological; S – Sexual; CTS – Conflict Tactics Scale; CADRI – Conflict in Adolescent Dating Relationships Inventory

[1]Location: USA

[2]Location: Canada

[3]Location: International (Israel, Mexico, South Africa, Spain, Thailand)

[a]Hand calculated

[b]Four items from the original CADRI, plus two additional items.

[c]Range for all items in category. For psychological aggression, this includes the range over the threatening behavior (range, 12.2-35.4%), relational aggression (range, 18.4-22.9%) and verbal/emotional abuse (range, 23.9-78.0%) sub-scales. Because of the high endorsement of verbal/emotional items, prevalence for this sub-scale represents scores above the median.

[d]Participants were involved with Child Protective Services, and were part of an intervention study. Prevalence is reported for control group participants (*n*=62) at their final assessment (timing of final assessment varied by participant). A portion of this sample was also used in the CADRI validation (partner agreement data; Wolfe et al., 2001).

[e]Range is emotional abuse – threatening behavior.

[f]Reported for dating violence victims (*n*=57) only.

[g]Mild and severe physical victimization. Severe scale contained two sexual violence items. Prevalence is calculated out of the total sample (*n*=197).

[h]13% experienced mild victimization and 23% experienced severe victimization (severe victimization includes the sexual items).

[i]Participants were enrolled in secondary school education (*M (SD)* age=16.08 (1.32), range 14-20).

[j]For the 378 participants with a partner in the last 6 months (53.2% female); 5.3% experienced frequent violence, and 32.6% experienced occasional violence.

[k]Racial/ethnic distribution was reported for participants' mothers, but not participants themselves. Mothers were 53.3% White, 36.0% Hispanic and 4.7% Black. Approximately half of the sample came from families were the mother had experienced intimate partner violence.

[l]Range of prevalence over all five items.

[m]The text reports 18 items, but only 13 items are listed in Table 1, where prevalence is reported.
[n]Mean age was 17.0 (range, 16-20), but all participants were drawn from high schools.
[o]Although the paper includes four psychological aggression items from the M-CTS, prevalence was not reported for these items.
[p]Sub-sample of Cascardi et al. (1999), which validated the M-CTS for use in adolescent populations. O'Leary et al. (2008) and O'Leary & Slep (2003) both use this sample. O'Leary & Slep (2003) use only control participants who were in the same relationship at baseline and post-test (~14 weeks later), and O'Leary et al. (2008) use all baseline data (treatment and control).
[q]Sears & Byers (2010) and Sears et al. (2007) use the same sample. Results are presented from Sears et al. (2007).
[r]This study randomly sampled adolescents aged 14 to 19 (*M* age ~16) in the Bangkok, Thailand area. Participants were recruited from high schools, vocational schools, and out-of-school settings.
[s]Of those who had dated (*n*=695). Range for psychological is threatening behaviors (49.2%), verbal/emotional abuse (49.2%) – relationally abusive behaviors (65.8%).
[t]Only the 778 participants who had dated are included in the analysis. Sherer (2009) and Sherer & Sherer (2008) use the same sample. Results are presented from Sherer & Sherer (2008).
[u]Of dating boys, 35.3% experienced relational aggression, 42.3% experienced threatening behavior and 88.9% experienced verbal aggression.
[v]Giordano et al. (2010) and Richards & Branch (2012) both use data from the Toledo Adolescent Relationship Study. Results are presented from Richards & Branch (2012).
[w]Participants were all in secondary school, but ages ranged from 13 to 23 (*M* age, boys=17.0; *M* age, girls=16.0). A minority of the sample was older than age 19 (5.8% of males and 1.8% of females).
[x]Data come from same sample as was used to validate the RFDE (Schultz and Jaycox, 2008). Prevalence number is for recent daters only (*n*=348).

Appendix 3.C
Male Perpetration Prevalence Estimates

Community-based samples									
							Perpetration Prevalence (%)		
Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CADRI									
Antônio & Hokoda (2009)[3]	CADRI – Spanish	14 (4 P, 10 Ps)	285	64.0	Mexican	Past year	4.9-13.6[a]	19.4-82.5[a]	
Archer et al. (2010)[3]	CADRI – Spanish	6 (P)[b]	601	58.2[a]	Spanish	In relationship	21.6		
Fernández-Fuertes & Fuertes (2010)[3]	CADRI – Spanish	14 (4 P, 10 Ps)	567	58.4[a]	Spanish	In a relationship (past year)	16.1	95.3	
Schiff & Zeira (2005)[3]	CADRI – Hebrew	25 (4 P, 17 Ps, 4 S)	105	51.0	Israeli	Past year	12.0-26.5[c]	10.2-80.0[c]	8.2-26.5[c]
Walton et al. (2009)	CADRI – modified	2 (P)	1128	54.1	Multi-racial (58.0% Black, 36.1% White)	Past year	8.3-9.7[d]		
Wolfe et al. (2003)[2]	CADRI	n/a (P)	158[e]	50.0[a]	Primarily White	Past 2 months (current or recent dating partner)	19.0		

(Appendix 3.C continues)

(Appendix 3.C continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Boafo (2011)[3]	CTS2	3 (2 P, 1 Ps)[f]	3747	48.2	South African	Ever	9.7-16.7	12.1	
Connolly et al. (2010)[2]	CTS2 – modified	9 (P)	664[g]	55.3[a]	Primarily White (73.7%; 9.2% Black; 12.1% Asian)	Past 6 months (current/ past dating relationship)	37.4		
Connolly et al. (2010)[3]	CTS2 – Italian, modified	9 (P)	578[g]	45.5[a]	Primarily White (95.8%; Italian)	Past 6 months (current/past dating relationship)	34.9		
Jiménez et al. (2008)[3]	CTS	9 (P)	446[h]	52.5	Spanish	Within the couple (past 6 months)	47.6[i]		
Lavoie et al. (2002)[2]	CTS[j]	7 (P)	717	0.0	French Canadian	Past year[k]	2.9-17.6[l]		
Lichter & McCloskey (2004) and McCloskey & Lichter (2003)[1]	CTS	6 (5 P, 1 S)	208	51.0	n/a[m]	Past year	0.9-12.1[n]		0.0

(Appendix 3.C continues)

(Appendix 3.C continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Muñoz-Rivas et al. (2007)[3]	M-CTS – Spanish	13 (8 P, 5 Ps)[o]	2416[p]	58.6	Spanish	Actual partner	31.7	92.8	
O'Leary & Slep (2003)[1]	M-CTS – Adolescent	10 (P)[q]	206[r]	58.3[a]	Primarily White (63.6%; 15.5% Black, 8.3% Hispanic)[a]	Duration of current dating relationship	17.4		
O'Leary et al. (2008)[1]	M-CTS – Adolescent	14 (8 P, 6 Ps)	2363[r]	49.8[a]	Multi-racial (55.1% White, 14.0% Black, 17.9% Hispanic)	Duration of current / most recent relationship	24.0	85.0	
Schumacher & Slep (2004)[1]	M-CTS – Adolescent	5 (Ps)	398[r]	57.0[a]	Multi-racial (56.9% White, 18.1% Black, 13.6% Hispanic)[a]	Duration of current / most recent relationship		84.0	
Sears et al. (2007)[2]	CTS – modified	7 (3 P, 4 Ps)	633	48.8[a]	Primarily White	Ever	15.0	35.0	

(Appendix 3.C continues)

(Appendix 3.C continued)

Probability/census samples									
Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Giordano et al. (2010) and Richards & Branch (2012)[1]	Revised CTS	4 (P)	970[s]	51.0[a]	Primarily White (64%; 23% Black, 11% Hispanic)	Current/ most recent dating partner	26.5[a]		
Swart et al. (2002)[3]	CTS2	20 (P)	928	53.2[a,t]	South African	Past year	35.3		

Abbreviations: P – Physical; Ps – Psychological; S – Sexual; CTS – Conflict Tactics Scale; CADRI – Conflict in Adolescent Dating Relationships Inventory

[1]Location: USA

[2]Location: Canada

[3]Location: International (Israel, Italy, Mexico, South Africa, Spain)

[a] Hand calculated

[b]Four items from the original CADRI, plus two additional items.

[c]Range for all items in category. For psychological aggression, this includes the range over the threatening behavior (range, 10.2-27.1%), relational aggression (range, 12.2-16.3%) and verbal/emotional abuse (range, 22.0-80.0%) sub-scales. Because of the high endorsement of verbal/emotional items, prevalence for this sub-scale represents scores above the median.

[d]Range is for moderate – severe aggression. Participants in this study were recruited from an emergency department (aged 14-18).

[e]Participants were involved with Child Protective Services, and were part of an intervention study. Prevalence is reported for control group participants (*n*=62) at their final assessment (timing of final assessment varied by participant). Prevalence of threatening behavior and emotional/verbal abuse was reported for victimization but not perpetration, and so is not presented in this table. A portion of this sample was also used in the CADRI validation (partner agreement data; Wolfe et al., 2001).

[f]While a sexual aggression item for perpetration was also included, it asked if the participant had ever forced anyone else to have sex with them when they did not want to (while the victimization item asked if a boy/girlfriend physically forced them to have sex when they did not want to). Thus, the prevalence of sexual perpetration (11.1%) is not included in this table, because it is not specific to dating violence.

[g]Uses CTS2 scale that was validated by Nocentini et al. (2011) , and also appears to use the same sample as Nocentini et al. (2011).

[h]Participants were enrolled in secondary school education (*M (SD)* age=16.08 (1.32), range 14-20).

[i]For the 378 participants with a partner in the last 6 months (53.2% female); 4.8% perpetrated frequent violence, and 42.7% perpetrated occasional violence.

[j]Sample was French-Canadian, so this instrument was likely translated, but this is not specified by authors.

[k]Represents any dating violence perpetrated during two years of data collection (data collected at age 16 and 17).

[l]Range represents perpetration of physical only and perpetration of physical and psychological, respectively. Psychological only prevalence not reported because it was collected using a non-validated scale.

[m]Racial/ethnic distribution was reported for participants' mothers, and not participants themselves. Mothers were 53.3% White, 36.0% Hispanic and 4.7% Black.

Approximately half of the sample came from families were the mother had experienced intimate partner violence. In McCloskey and Lichter (2003), the racial/ethnic distribution of the children was described as 53.7% White, 35.8% Hispanic and 4.7% Black. Lichter & McCloskey (2004) and McCloskey & Lichter (2003) use the same sample; results from Lichter & McCloskey (2004) are reported.

[n]Range of prevalence over all five items. In McCloskey and Lichter (2003), the prevalence of any physical aggression perpetration was listed as 11.1% (*n*=296).

[o]The text reports 18 items, but only 13 items are listed in Table 1, where prevalence is reported.

[p]Mean age was 17.0 (range, 16-20). All participants were drawn from high schools.

[q]Although the paper includes four psychological aggression items from the M-CTS, prevalence was not reported for these items.

[r]Sub-sample of Cascardi et al. (1999), which validated the M-CTS for use in adolescent populations. O'Leary et al. (2008), O'Leary & Slep (2003) and Schumacher & Slep (2004) all use this sample. O'Leary & Slep (2003) use only control participants who were in the same relationship at baseline and post-test (~14 weeks later), O'Leary et al. (2008) use all baseline data (treatment and control), and Schumacher & Slep (2004) use control participants who had dated during the study period, and who were assessed at baseline and 3-month follow-up.

[s]Giordano et al. (2010) and Richards & Branch (2012) both use data from the Toledo Adolescent Relationship Study. Results are presented from Richards & Branch (2012).

[t]Participants were all in secondary school, but ages ranged from 13 to 23 (*M* age, boys=17.0; *M* age, girls=16.0). A minority of the sample was older than age 19 (5.8% of males and 1.8% of females).

Appendix 3.D

Female Victimization Prevalence Estimates

Community-based samples									
							Victimization Prevalence (%)		
Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CADRI									
Archer et al. (2010)[3]	CADRI – Spanish	6 (P)[b]	601	58.2[a]	Spanish	In relationship	23.0		
Fernández-Fuertes & Fuertes (2010)[3]	CADRI – Spanish	14 (4 P, 10 Ps)	567	58.4[a]	Spanish	In a relationship (past year)	17.5	95.5	
Schiff & Zeira (2005)[3]	CADRI – Hebrew	25 (4 P, 17 Ps, 4 S)	105	51.0	Israeli	Past year	9.6-13.7[c]	5.8-84.9[c]	1.9-39.6[c]
Wolfe et al. (2003)[2]	CADRI	n/a (P, Ps)	158[d]	50.0[a]	Primarily White	Past 2 months (current or recent dating partner)	18.0	27.0-32.0[e]	
CTS									
Black et al. (2008)	CTS2 – modified	14 (P, S)	197	56.0[f]	Multi-racial (47% Black, 41% White)[f]	n/a	16.2[a,g]		
Boafo (2011)[3]	CTS2	5 (3 P, 1 Ps, 1 S)	3747	48.2	South African	Ever	2.7-11.1[h]	4.2	5.1

(Appendix 3.D continues)

(Appendix 3.D continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Buzy et al. (2004)[1]	CTS2	15 (12 P, 3 S – modified)	106	100	Multi-racial (69% Hispanic, 29% Black)	Past 4 months (most recent boyfriend)	20.0[i]		2.0[i]
Callahan et al. (2003)[1]	CTS2	14 (12 P – modified, 2 S)	190	47.4[a]	Multi-racial (50.5% Black, 49.5% White)[a]	Ever	57.0[j]		
Collin-Vézina et al. (2006) and Manseau et al. (2008)[2]	CTS2 – French (short)	33 (12 P, 8 Ps, 7 S, 6 Inj)	196[k]	100	French Canadian	Ever	67.2[l]	87.9[l]	70.2[l]
Cyr et al. (2006)[2]	CTS2 – French	12 (P)	126	100	French Canadian	Past year	45.2[m]		
Jiménez et al. (2008)[3]	CTS	9 (P)	446[n]	52.5	Spanish	Within the couple (past 6 months)	42.8[o]		
Lichter & McCloskey (2004)[1]	CTS	6 (5 P, 1 S)	208	51.0	n/a[p]	Past year	2.9-24.8[q]		8.1
Muñoz-Rivas et al. (2007)[3]	M-CTS – Spanish	13 (8 P, 5 Ps)[r]	2416[s]	58.6	Spanish	Actual partner	37.4	93.7	

(Appendix 3.D continues)

(Appendix 3.D continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
O'Leary & Slep (2003)[1]	M-CTS – Adolescent	10 (P)[t]	206[u]	58.3[a]	Primarily White (63.6%; 15.5% Black, 8.3% Hispanic)[a]	Duration of current dating relationship	22.3		
O'Leary et al. (2008)[1]	M-CTS – Adolescent	14 (8 P, 6 Ps)	2363[u]	49.8[a]	Multi-racial (55.1% White, 14.0% Black, 17.9% Hispanic)	Duration of current / most recent relationship	30.0	88.0	
Sears et al. (2007) and Sears & Byers (2010)[2]	CTS – modified	7 (3 P, 4 Ps)	633[v]	48.8[a]	Primarily White	Ever	29.0	61.0	
Other									
Gagné et al. (2005)[2]	VIFFA	40 (17 P, 19 Ps, 4 S)	622	100	Québec or Canadian culture (79%)	Past year	25.0-29.0[w]	37.0	26.0
Probability/census samples									
CADRI									
Pradubmook-Sherer (2009)[3]	CADRI – Thai	23 (4 P, 15 Ps, 4 S)	1296[x]	54.5	Thai	n/a	41.2[y]	46.7-59.0[y]	46.7[y]

(Appendix 3.D continues)

(Appendix 3.D continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CADRI									
Sherer (2009) and Sherer & Sherer (2008)[3]	CADRI – Hebrew & Arabic	23 (4 P, 15 Ps, 4 S)	1357[z]	56.4[a]	Israeli (47.2% Jewish, 52.8% Arab)	Current/ most recent partner	32.8	13.3-86.7[aa]	21.7
CTS									
Giordano et al. (2010) and Richards & Branch (2012)[1]	Revised CTS	4 (P)	970[bb]	51.0[a]	Primarily White (64%; 23% Black, 11% Hispanic)	Current/ most recent dating partner	13.9[a]		
Swart et al. (2002)[3]	CTS2	20 (P)	928	53.2[a,cc]	South African	Past year	41.7		
Vézina et al. (2011)[2]	CTS – French	7 (P)	550	100	French Canadian	Past year (most difficult relationship)	9.6		
Other									
Ulloa et al. (2004)[1]	RFDE	11	678[dd]	55.0	100% Hispanic	Past 6 months		55.2	

Abbreviations: P – Physical; Ps – Psychological; S – Sexual; Inj – Injury; CTS – Conflict Tactics Scale; CADRI – Conflict in Adolescent Dating Relationships Inventory

[1]Location: USA

[2]Location: Canada

[3]Location: International (Israel, South Africa, Spain, Thailand)

[a]Hand calculated

[b]Four items from the original CADRI, plus two additional items.

[c]Range for all items in category. For psychological aggression, this includes the range over the threatening behavior (range, 9.4-23.5%), relational aggression (range, 5.8-18.0%) and verbal/emotional abuse (range, 15.7-84.9%) sub-scales. Because of the high endorsement of verbal/emotional items, prevalence for this sub-scale represents scores above the median.

[d]Participants were involved with Child Protective Services (CPS), and were part of an intervention study. Prevalence is reported for control group participants (*n*=62) at their final assessment (timing of final assessment varied by participant). A portion of this sample was also used in the CADRI validation (partner agreement data; Wolfe et al., 2001).
[e]Range is threatening behavior – emotional abuse.
[f]Reported for dating violence victims (*n*=57) only.
[g]Mild and severe physical victimization. Severe scale contained two sexual violence items. Prevalence is calculated out of the total sample (*n*=197).
[h]Range over the three items.
[i]For participants with a boyfriend at Time 1 or Time 2 - 20% reported only physical violence, 2% reported only sexual violence and 14% reported both physical and sexual acts.
[j]18% experienced mild victimization and 39% experienced severe victimization (severe victimization includes the sexual items).
[k]Sample was comprised of adolescents living in a Child Protective Services center (living there because of child maltreatment, or because of their own delinquency). Collin-Vézina et al. (2006) and Manseau et al. (2008) use the same sample; results from Manseau et al. (2008) are reported here.
[l]Represents any victimization of that type. For psychological aggression, 58.3% experienced severe aggression and 85.6% experienced minor aggression. For physical aggression, 53.1% experienced severe aggression and 60.9% experienced minor aggression. For sexual aggression, 35.1% experienced severe aggression and 67.2% experienced minor aggression. 44.0% of the sample reported injuries following victimization (26.8% severe, 40.5% minor).
[m]Sample was comprised of adolescents who were referred to CPS for alleged sexual abuse. In total, 45.2% experienced physical violence victimization by a dating partner, with 43.7% experiencing minor violence, and 20.6% experiencing severe violence.
[n]Participants were enrolled in secondary school education (*M (SD)* age=16.08 (1.32), range 14-20).
[o]For the 378 participants with a partner in the last 6 months (53.2% female); 3.6% experienced frequent violence, and 39.2% experienced occasional violence.
[p]Racial/ethnic distribution was reported for participants' mothers, and not participants themselves. Mothers were 53.3% White, 36.0% Hispanic and 4.7% Black. Approximately half of the sample came from families were the mother had experienced intimate partner violence.
[q]Range of prevalence over all five items.
[r]The text reports 18 items, but only 13 items are listed in Table 1, where prevalence is reported.
[s]Mean age was 17.0 (range, 16-20), but all participants were drawn from high schools.
[t]Although the paper includes four psychological aggression items from the M-CTS, prevalence was not reported for these items.
[u]Sub-sample of Cascardi et al. (1999), which validated the M-CTS for use in adolescent populations. O'Leary et al. (2008) and O'Leary & Slep (2003) both use this sample. O'Leary & Slep (2003) use only control participants who were in the same relationship at baseline and post-test (~14 weeks later), and O'Leary et al. (2008) use all baseline data (treatment and control).
[v]Sears & Byers (2010) and Sears et al. (2007) use the same sample. Results are presented from Sears et al. (2007).
[w]Range is indirect (e.g., broke an object belonging to you on purpose, in a moment of anger or frustration) and direct physical violence victimization, respectively. Participants in this study were aged 14-20 (*M (SD)*=16.3(0.8)), but were all enrolled in the 10th and 11th grades.
[x]This study randomly sampled adolescents aged 14 to 19 (*M* age ~16) in the Bangkok, Thailand area. Participants were recruited from high schools, vocational schools, and out-of-school settings.
[y]Of those who had dated (*n*=695). Range for psychological is threatening behaviors (46.7%), verbal/emotional abuse (46.7%) – relationally abusive behaviors (59.0%).
[z]Only the 778 participants who had dated are included in the analysis. Sherer (2009) and Sherer & Sherer (2008) use the same sample. Results are presented from Sherer & Sherer (2008).
[aa]Of girls, 13.3% experienced relational aggression, 26.4% experienced threatening behavior, and 86.7% experienced verbal/emotional aggression.
[bb]Giordano et al. (2010) and Richards & Branch (2012) both use data from the Toledo Adolescent Relationship Study. Results are presented from Richards & Branch (2012).
[cc]Participants were all in secondary school, but ages ranged from 13 to 23 (*M* age, boys=17.0; *M* age, girls=16.0). A minority of the sample was older than age 19 (5.8% of males and 1.8% of females).

[dd]Data come from same sample as was used to validate the RFDE (Schultz and Jaycox, 2008). Prevalence number is for recent daters only (*n*=348).

Appendix 3.E

Female Perpetration Prevalence Estimates

Community-based samples									
							Perpetration Prevalence (%)		
Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CADRI									
Antônio & Hokoda (2009)[3]	CADRI – Spanish	14 (4 P, 10 Ps)	285	64.0	Mexican	Past year	15.4-16.5[a]	20.9-79.1[a]	
Archer et al. (2010)[3]	CADRI – Spanish	6 (P)[b]	601	58.2[a]	Spanish	In relationship	28.6		
Fernández-Fuertes & Fuertes (2010)[3]	CADRI – Spanish	14 (4 P, 10 Ps)	567	58.4[a]	Spanish	In a relationship (past year)	30.2	97.0	
Schiff & Zeira (2005)[3]	CADRI – Hebrew	25 (4 P, 17 Ps, 4 S)	105	51.0	Israeli	Past year	11.3-38.5[c]	3.8-94.2[c]	0.0-28.3[c]
Walton et al. (2009)[1]	CADRI – modified	2 (P)	1128	54.1	Multi-racial (58.0% Black, 36.1% White)	Past year	15.6-20.2[d]		
Wolfe et al. (2003)[2]	CADRI	n/a (P)	158[e]	50.0[a]	Primarily White	Past 2 months (current or recent dating partner)	41.0		

(Appendix 3.E continues)

(Appendix 3.E continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Boafo (2011)[3]	CTS2	3 (2 P, 1 Ps)[f]	3747	48.2	South African	Ever	2.4-10.4	2.7	
Connolly et al. (2010)[2]	CTS2 – modified[g]	9 (P)	664	55.3[a]	Primarily White (73.7%; 9.2% Black, 12.1% Asian)	Past 6 months (current/past dating relationship)	28.6		
Connolly et al. (2010)[3]	CTS2 – Italian, modified[g]	9 (P)	578	45.5[a]	Primarily Italian (95.8%)	Past 6 months (current/past dating relationship)	31.9		
Jiménez et al. (2008)[3]	CTS	9 (P)	446[h]	52.5	Spanish	Within the couple (past 6 months)	51.9[i]		
Kernsmith & Tolman (2011)[1]	CTS2 – modified	10 (9 P, 1 S)	102	100	Multi-racial (50.5% Black, 44.1% White)	Worst experience with dating violence perpetration	28.0-59.0[j]		
Lichter & McCloskey (2004) and McCloskey & Lichter (2003)[1]	CTS	5 (P)	208	51.0	n/a[k]	Past year	2.9-18.3[l]		

(Appendix 3.E continues)

(Appendix 3.E continued)

Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Muñoz-Rivas et al. (2007)[3]	MCTS – Spanish	13 (8 P, 5 Ps)[m]	2416[n]	58.6	Spanish	Actual partner	41.9	95.3	
O'Leary & Slep (2003)[1]	MCTS – Adolescent	10 (P)[o]	206[p]	58.3[a]	Primarily White (63.6%; 15.5% Black, 8.3% Hispanic)[a]	Duration of current dating relationship	30.6		
O'Leary et al. (2008)[1]	MCTS – Adolescent	14 (8 P, 6 Ps)	2363[p]	49.8[a]	Multi-racial (55.1% White, 14.0% Black, 17.9% Hispanic)	Duration of current / most recent relationship	40.0	92.0	
Schumacher & Slep (2004)[1]	MCTS – Adolescent	5 (Ps)	398[p]	57.0[a]	Multi-racial (56.9% White, 18.1% Black, 13.6% Hispanic)[a]	Duration of current / most recent relationship		94.0	
Sears et al. (2007)[2]	CTS – modified	7 (3 P, 4 Ps)	633	48.8[a]	Primarily White	Ever	28.0	47.0	

(Appendix 3.E continues)

(Appendix 3.E continued)

Probability/census samples									
Authors and year	*Measure*	*Number of items*	*N*	*% female*	*Race/ ethnicity*	*Time frame*	*P*	*Ps*	*S*
CTS									
Giordano et al. (2010) and Richards & Branch (2012)[1]	Revised CTS	4 (P)	970[q]	51.0[a]	Primarily White (64%; 23% Black, 11% Hispanic)	Current/ most recent dating partner	19.4[a]		
Swart et al. (2002)[3]	CTS2	20 (P)	928[r]	53.2[a]	South African	Past year	43.5		

Abbreviations: P – Physical; Ps – Psychological; S – Sexual; CTS – Conflict Tactics Scale; CADRI – Conflict in Adolescent Dating Relationships Inventory
[1]Location: USA
[2]Location: Canada
[3]Location: International (Israel, Italy, Mexico, South Africa, Spain)

[a] Hand calculated
[b]Four items from the original CADRI, plus two additional items.
[c]Range for all items in category. For psychological aggression, this includes the range over the threatening behavior (range, 9.4-38.5%), relational aggression (range, 3.8-5.8%) and verbal/emotional abuse (range, 17.0-94.2%) sub-scales. Because of the high endorsement of verbal/emotional items, prevalence for this sub-scale represents scores above the median.
[d]Range is severe – moderate. Participants in this study were recruited from an emergency department (aged 14-18).
[e]Participants were involved with Child Protective Services, and were part of an intervention study. Prevalence is reported for control group participants (*n*=62) at their final assessment (timing of final assessment varied by participant). Prevalence of threatening behavior and emotional/verbal abuse was reported for victimization but not perpetration, and so is not presented in this table. A portion of this sample was also used in the CADRI validation (partner agreement data; Wolfe et al., 2001).
[f]While a sexual aggression item for perpetration was also included, it asked if the participant had ever forced anyone else to have sex with them when they did not want to (while the victimization item asked if a boy/girlfriend physically forced them to have sex when they did not want to). Thus, the prevalence of sexual perpetration (1.5%) is not included in this table, because it is not specific to dating violence.
[g]Uses CTS2 scale that was validated by Nocentini et al. (2011). Also appears to be the same sample as Nocentini et al. (2011).
[h]Participants were enrolled in secondary school education (*M (SD)* age=16.08 (1.32), range 14-20).
[i]For the 378 participants with a partner in the last 6 months (53.2% female); 4.4% perpetrated frequent violence, and 47.5% perpetrated occasional violence.
[j]Range reflects use of severe and mild violence, respectively. Severe violence includes the sexual violence item (only one participant reported perpetration of sexual violence). Low SES sample.
[k]Racial/ethnic distribution was reported for participants' mothers, and not participants themselves. Mothers were 53.3% White, 36.0% Hispanic and 4.7% Black. Approximately half of the sample came from families were the mother had experienced intimate partner violence. In McCloskey and Lichter (2003), the racial/ethnic distribution of the children was described as 53.7% White, 35.8% Hispanic and 4.7% Black. Lichter & McCloskey (2004) and McCloskey & Lichter

(2003) use the same sample; results from Lichter & McCloskey (2004) are reported.
[l]Range of prevalence for all five items. This same sample was examined in McCloskey and Lichter (2003), where prevalence of any physical aggression perpetration was 20.3% (*n*=296).
[m]The text reports 18 items, but only 13 items are listed in Table 1, where prevalence is reported.
[n]Mean age was 17.0 (range, 16-20). All participants were drawn from high schools.
[o]Although the paper includes four psychological aggression items from the M-CTS, prevalence was not reported for these items.
[p]Sub-sample of Cascardi et al. (1999), which validated the M-CTS for use in adolescent populations. O'Leary et al. (2008), O'Leary & Slep (2003) and Schumacher & Slep (2004) all use this sample. O'Leary & Slep (2003) use only control participants who were in the same relationship at baseline and post-test (~14 weeks later), O'Leary et al. (2008) use all baseline data (treatment and control), and Schumacher & Slep (2004) use control participants who had dated during the study period, and who were assessed at baseline and 3-month follow-up.
[q]Giordano et al. (2010) and Richards & Branch (2012) both use data from the Toledo Adolescent Relationship Study. Results are presented from Richards & Branch (2012).
[r]Participants were all in secondary school, but ages ranged from 13 to 23 (*M* age, boys=17.0; *M* age, girls=16.0). A minority of the sample was older than age 19 (5.8% of males and 1.8% of females).

Appendix 3.F
Measure Quality Assessment Template

Measurement Quality Assessment

~Explain all assessments~

Adapted from:

Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure, and results. *International Journal of Testing, 10*, 295-317.

Scientific Advisory Committee of Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11*, 193-205.

Shields, A., Gwaltney, C., Paty, J., & Shiffman, S. (2006). *Documentation of PRO instruments to meet contemporary FDA standards*. Retrieved from http://www.invivodata.com/media/files/PRO%20Consulting_White%20Paper_Documentation%20of%20PRO%20Instruments.pdf

Zaza, S. Wright-De Agüero, L. K., Briss, P. A., Truman, B. I., Hopkins, D. I., Hennessy, M. H.,…Task Force on Community Preventive Services. (2000). Data collection instrument and procedure for systematic reviews in the *Guide to Community Preventive Services*. *American Journal of Preventive Medicine, 18*, 44-74.

1. Descriptions

a) Was the intended population well described (including age and gender)? ☐ Yes ☐ No

b) Was the specific study population well described? ☐ Yes ☐ No

c) Are test items free from racist, ethnocentric, or sexist content or any other content offensive to specific groups of people? ☐ Yes ☐ No

d) Was the measure development process well described?

i. Concept to be measured ☐ Yes ☐ No

ii. Conceptual/empirical/theoretical bases for item content ☐ Yes ☐ No

iii. Target population involvement in content derivation ☐ Yes ☐ No

iv. Information on distinctiveness of scales ☐ Yes ☐ No

v. Evidence of scale variability ☐ Yes ☐ No

vi. Intended level of measurement ☐ Yes ☐ No

vii. Rationale for deriving scale scores ☐ Yes ☐ No

e) Was the measure well described?

i. Format List:

ii. Are the instructions for the subject of the test clear and complete? ☐ Yes ☐ No

iii. Purpose of test? ☐ Yes ☐ No

iv. Number of items

v. Are the test items standardized? ☐ Yes ☐ No

vi. Recall period List:

vii. Response options (number and type) List:

viii. Data collection method List:

ix. Are the items, test booklet, answering scales, and answer form devised in a way that errors can be avoided when filling in? ☐ Yes ☐ No

x. Scoring procedures (computerized or objective system?) List:

xi. Is the scoring system devised in such a way that errors can be avoided during scoring? ☐ Yes ☐ No

*Explain:*_____

2. Sampling

a)	Did the authors specify the sampling frame or universe of selection for the study population?	☐ Yes	☐ No	☐ N/A
b)	Did the authors specify the criteria for study eligibility?	☐ Yes	☐ No	☐ N/A
c)	Was the population that served as the unit of analysis the entire eligible population or a probability sample at the point of observation?	☐ Yes	☐ No	☐ N/A
d)	Are there selection bias issues not otherwise addressed?	☐ Yes	☐ No	☐ N/A

*Explain:*_____

3. Reliability: Were the outcomes variables reliable?

a)	Are the findings of the reliability research sufficient with respect to the intended type of decisions to be made with the aid of the test?	☐ Yes	☐ No	☐ N/A
	i. Parallel-form reliability	☐ Yes	☐ No	☐ N/A
	ii. Reliability based on item inter-relations	☐ Yes	☐ No	☐ N/A
	iii. Test-retest reliability	☐ Yes	☐ No	☐ N/A
	iv. Inter-rater reliability	☐ Yes	☐ No	☐ N/A
	v. Methods based on item-response theory	☐ Yes	☐ No	☐ N/A
	vi. Methods based on generalizability theory or SEM	☐ Yes	☐ No	☐ N/A
b)	Reliability estimates and standard errors for all score elements (classic test) **OR** standard error of the mean over the range of scale and marginal reliability of each scale (modern IRT)	☐ Yes	☐ No	☐ N/A
c)	Data to calculate reliability coefficients or actual calculations of reliability coefficients	☐ Yes	☐ No	☐ N/A
d)	Are procedures for computing reliability coefficients correct?	☐ Yes	☐ No	☐ N/A
e)	Are samples for computing the reliability coefficients consistent with the intended use of the test?	☐ Yes	☐ No	☐ N/A
f)	Well-argued rationale to support design of study and the interval between first and subsequent administration to support the assumption that the population is stable?	☐ Yes	☐ No	☐ N/A
g)	Information on test-retest reliability and inter-rater reliability based on intraclass correlation coefficients	☐ Yes	☐ No	☐ N/A
h)	Information on the comparability of the item parameter estimates and on measurement precision over repeated administrations	☐ Yes	☐ No	☐ N/A
i)	Reliability data for each major population of interest, if applicable	☐ Yes	☐ No	☐ N/A
j)	Is it possible to make a thorough judgment of the reliability of the test on the basis of the information given?	☐ Yes	☐ No	☐ N/A

*Explain:*_____

4. Validity

a) Were the outcome variables valid? (general)

i.	Rationale supporting the particular mix of evidence presented for intended uses	Describe below		
ii.	Clear description of the methods employed to collect validity data	☐ Yes	☐ No	☐ N/A
iii.	Detailed description of the sample used to examine validity (see section 2)	☐ Yes	☐ No	☐ N/A
iv.	Validity data presented for each major population of interest	☐ Yes	☐ No	☐ N/A
v.	Hypotheses tested and data relating to tests	Describe below		
vi.	Clear rationale and support for choice of criteria measures	☐ Yes	☐ No	☐ N/A

b) Is there information provided about the construct validity of the test? ☐ Yes ☐ No ☐ N/A

If yes, do the findings of the validity research support the intended construct(s) being measured—or do the findings of the validity research make clear what is being measured—on the basis of information on:

i.	The dimensionality of the scores?	☐ Yes	☐ No	☐ N/A
ii.	The psychometric quality of the items?	☐ Yes	☐ No	☐ N/A
iii.	The invariance of the factor structure and potential item bias across subgroups	☐ Yes	☐ No	☐ N/A
iv.	Convergent and discriminant validity	☐ Yes	☐ No	☐ N/A
v.	Differences between relevant groups	☐ Yes	☐ No	☐ N/A
vi.	Other findings	☐ Yes	☐ No	☐ N/A

What is the quality of the construct validity research?

vii.	Are the procedures used in obtaining and computing data on construct validity correct?	☐ Yes	☐ No	☐ N/A
viii.	Are the samples used in the research on construct validity consistent with groups for whom the test is intended?	☐ Yes	☐ No	☐ N/A

ix.	What is the quality of the other measures used in the construct validity research?	Describe below		
x.	Is the quality of the research, as rated in items vii to ix, good enough to corroborate the assessment of the construct validity as given in items i-vi?	☐ Yes	☐ No	☐ N/A
c)	Is there information about the test-criterion relationship?	☐ Yes	☐ No	☐ N/A
d)	If yes, are the findings of the validity research sufficient with respect to the type of decisions to be made with the test?	☐ Yes	☐ No	☐ N/A

What is the quality of the criterion validity research?

i.	Are the procedures used in collecting and calculating data on criterion validity correct?	☐ Yes	☐ No	☐ N/A
ii.	Are the samples used in the research on criterion validity consistent with the intended use of the test?	☐ Yes	☐ No	☐ N/A
iii.	What is the quality of criterion measures?	Describe below		
iv.	Is the quality of the research, as rated in items i to iii, good enough to corroborate the assessment of the criterion validity as given in d)?	☐ Yes	☐ No	☐ N/A

*Explain:*_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

5. Data Analysis: *Did the authors conduct appropriate statistical testing by…*

a)	Conducting statistical testing (where appropriate)?	☐ Yes	☐ No	☐ N/A
b)	Reporting which statistical tests were used?	☐ Yes	☐ No	☐ N/A
c)	Are there problems with the data analysis?	☐ Yes	☐ No	☐ N/A

*Explain:*_____

6. Interpretation of Results

a)	Was there at least a 30% response rate?	☐ Yes	☐ No	☐ N/A
b)	If longitudinal, did at least 80% of enrolled participants complete the study?	☐ Yes	☐ No	☐ N/A
c)	Rationale given for selection of external criteria of populations for purposes of comparison and interpretability of data	☐ Yes	☐ No	☐ N/A
d)	Information regarding the ways in which data from the instrument should be reported and displayed	☐ Yes	☐ No	☐ N/A
e)	Meaningful 'benchmarks' or norms to facilitate interpretation of scores	☐ Yes	☐ No	☐ N/A
f)	Is there information on possible differences between subgroups (e.g., due to gender or ethnicity)?	☐ Yes	☐ No	☐ N/A
g)	If domain-referenced interpretation:			
	i. Is there sufficient agreement between raters?	☐ Yes	☐ No	☐ N/A
	ii. Are the procedures for determining cut scores correct?	☐ Yes	☐ No	☐ N/A
	iii. Have the raters been selected and trained appropriately?	☐ Yes	☐ No	☐ N/A
h)	If criterion-referenced interpretation:			
	iv. Do the research findings justify the use of cut scores?	☐ Yes	☐ No	☐ N/A
	v. Is the composition of the research group consistent with the intended purpose of the test?	☐ Yes	☐ No	☐ N/A
	vi. Is the size of the research group sufficient?	☐ Yes	☐ No	☐ N/A

*Explain:*_____

_____

_____

_____

_____

7. Burden: *Was information provided on…*

a)	The average time and range of time needed to complete the instrument	☐ Yes	☐ No	☐ N/A
b)	Reading and comprehension level	☐ Yes	☐ No	☐ N/A
c)	Any special requirements or requests made of respondent	☐ Yes	☐ No	☐ N/A
d)	Evidence that the instrument places no undue physical or emotional strain on the respondent	☐ Yes	☐ No	☐ N/A
e)	When or under what circumstances the instrument is not suitable for respondents	☐ Yes	☐ No	☐ N/A
f)	Any resources required for administration of the instrument (if yes, explain below)	☐ Yes	☐ No	☐ N/A
g)	Average time and range of time required of a trained interviewer to administer the instrument in face-to-face interviews, by telephone, or with computer-assisted formats	☐ Yes	☐ No	☐ N/A
h)	Amount of training and level of education or professional expertise and experience needed by administrative staff	☐ Yes	☐ No	☐ N/A
i)	Instructions for test administrator (and are they clear and complete?)	☐ Yes	☐ No	☐ N/A

*Explain:*_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

8. Longitudinal Studies: *Was the study longitudinal?* ☐ Yes ☐ No

If yes...

a) Was there evidence of changes in scores of the instrument over time? ☐ Yes ☐ No

b) Was there data comparing a group that was expected to change with a group that was expected to remain stable? ☐ Yes ☐ No

c) Is the stable group of a sufficient size? ☐ Yes ☐ No

d) Is the stable group representative for the referred group(s)? ☐ Yes ☐ No

e) Population(s) on which responsiveness has been tested, including time intervals of assessment, the interventions or measures involved in evaluating change, and the populations assumed to be stable — List below

*Explain:*_____

9. Cultural and language adaptations: *Has the measure been adapted or translated?* ☐ Yes ☐ No

If yes…

a) Were methods used to achieve conceptual equivalence described? ☐ Yes ☐ No

b) Were methods used to achieve linguistic equivalence described? ☐ Yes ☐ No

c) Were there any significant differences between the original and translated versions? ☐ Yes ☐ No

d) If 'yes' to c), how were any inconsistencies reconciled? List below

*Explain:*_____

10. Other

Other important limitations of the study **not** identified elsewhere (specify):

If there are alternative modes of administration (e.g., self-report, interviewer-administered), is evidence on reliability, validity, responsiveness, interpretability, and burden provided for **each** mode of administration?

Appendix 4.A
Measurement Quality Assessment: Attitude Scales

Authors and year	Content	Sampling	Reliability	Validity
Attitudes Towards Dating Violence				
Hokoda et al. (2006)	Attitudes towards male and female use of dating violence	High school students (non-representative)	Internal consistency Test-retest	n/a
Price et al. (1999)	Attitudes towards male and female use of dating violence	Classrooms within three school districts (urban and rural; non-representative)	Internal consistency	Criterion Construct (convergent)
AADS/JVCT				
Muñoz-Rivas et al. (2011)	Attitudes justifying use of physical aggression (AADS) and attitudes justifying use of psychological aggression (JVCT)	High school students from 20 schools in Madrid (representative; schools were randomly selected)	Internal consistency	Construct (factor structure, convergent)
Slep et al. (2001)	Attitudes justifying use of physical aggression (AADS) and attitudes justifying use of psychological aggression (JVCT)	Students from seven high schools enrolled in mandatory health education classes (non-representative)	Internal consistency Test-retest	Concurrent Construct (factor structure, convergent)
Other – Attitudes				
Davidson (2005)	Attitudes about dating relationships (healthy and verbal, physical and sexual coercion)	St 1: Local adolescent organization St 2 and St 3: One private high school, one public high school St 4: One public high school (all non-representative)	Internal consistency Test-retest	Concurrent Construct (discriminant)

(Appendix 4.A continues)

(Appendix 4.A continued)

Authors and year	Content	Sampling	Reliability	Validity
Other – Attitudes				
Edelen et al. (2008) and Orlando et al. (2006)	Prescribed norms about dating violence and approval of retaliation	9th grade health classes serving primarily Latino populations (representative)	Internal consistency	Differential item functioning
Rayburn et al. (2007)	Think-aloud cognitive assessment paradigm, assessing reactions to audio-taped dating violence scenarios	Public high school students from 13 classrooms (non-representative)	Inter-rater Manipulation check	Participant debrief Construct
Slep et al. (2001)	Attitudes justifying physical aggression (AIV)	Students from seven high schools enrolled in mandatory health education classes (non-representative)	Internal consistency Test-retest	Construct (convergent)

Note. Abbreviations: AADS/JVCT – Attitudes About Aggression in Dating Situations / Justification of Verbal/Coercive Tactics; AIV – Attitudes Towards Interpersonal Violence; St – Study

Appendix 5.A
Initial PATS Item Pool

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Sadistic	
1A	Harmed or destroyed your personal things of value (e.g., pictures, keepsakes, clothes, etc.) as a way to intimidate you	1. Harmed or destroyed your personal things of value (e.g., pictures, clothing, cell phone) as a way to intimidate you[1] 2. Threatened to break or destroy something that was important to you / that you valued as a way to intimidate you[2]
1B	Threatened to harm others (e.g., your family, your close friends) around you to intimidate you[a]	3. Threatened to hurt one of your friends (physically or by bullying) in order to intimidate you[1] 4. Threatened to spread rumors about a friend in order to intimidate you[1] 5. Threatened to have his/her friends physically hurt someone you care about in order to intimidate you[2]
1C	Harmed pets as a way to intimidate you[a]	6. Threatened to hurt something that means a lot to you (for example, a pet) in order to intimidate you[1] 7. Drove recklessly when you were in the car in order to frighten you[2] 8. Pressured you to drink alcohol, take drugs, or do something illegal, when you did not want to[2]
	Threats	
2A	Threw a temper tantrum (e.g., breaking objects, acting in a rage) as a way to frighten you	1. Threw a temper tantrum (e.g., breaking/throwing objects, acting in a rage) as a way to frighten you[1] 2. Threw things at you (e.g., a cell phone) as a way to frighten you[1] 3. Continued to do something that scared you, even after you'd asked him/her to stop[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Threats	
2B	Verbally threatened to physically harm you or made a gesture that seemed physically threatening as a way to frighten you	4. Verbally threatened to physically harm you or made a gesture that seemed physically threatening as a way to frighten you[3] 5. Physically grabbed you to show that he/she was more powerful, as a way to scare you[1] 6. Said things to scare you (e.g., told you something "bad" would happen, threatened to commit suicide)[2] 7. Threatened you with a knife, gun or other weapon[2]
2C	Threatened to kill you as a way to frighten you[a]	8. **Threatened** to kill you as a way to frighten you (whether or not you thought he/she would go through with it)[1] 9. **Threatened** to kill you when you were planning to break up with him/her, or because he/she was scared of losing you[1]
	Isolating	
3A	Acted rude toward, gossip about, or tell lies about your family and friends to discourage you from spending time with them?	1. Acted rude toward, gossiped about, or told lies about your friends to discourage you from spending time with them?[1] 2. Was constantly physically affectionate, when you did not want him/her to be, in order to show his/her ownership of you[1] 3. Sent mean texts to your friends from your phone, or responded rudely to private Facebook or email messages from your account, in order to make people mad at you[1]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Isolating	
3B	Tried to keep you from socializing with family or friends without him/her being present?[a]	4. Tried to keep you from socializing with your friends without him/her being present, in order to isolate you?[1] 5. At school, tried to keep you away from everyone else, in order to isolate you?[1] 6. Tried to control your social life (who you talk to, what you do on your phone or the internet) in order to isolate you[1]
3C	Tried to forbid you from socializing with family or friends?[a]	7. Tried to forbid you from seeing your friends, in order to isolate you?[1] 8. Tried to forbid you from attending activities or events where he/she wouldn't be with you, in order to isolate you?[1] 9. Forced you to skip school so that he/she could keep you away from other people[1] 10. Pressured you to drop out of school so that he/she could isolate you[1] 11. Kept you from attending your after-school activities, or pressured you to drop out of your after-school activities, in order to isolate you[1]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
Isolating – Electronic Isolation		
n/a	n/a	12. Made you delete or de-activate a social media account (e.g., your Facebook account) in order to isolate you[1] 13. Told you that you were only allowed to use social media (e.g, Facebook, Twitter) when he/she was present[1] 14. Took your phone, and only let you use it when he/she was with you[1] 15. Forbid you from using your phone when you were with him/her, so that you couldn't communicate with anyone else[1] 16. Blocked people from your social media account(s), or deleted things from your account, when you did not want him/her to[1]
Manipulation		
4A	Continued to act very upset (e.g., pouted, stayed angry, gave you the silent treatment) until you did what he/she wanted you to do	1. Continued to act very upset (e.g., pouted, stayed angry, gave you the silent treatment) until you did what he/she wanted you to do[3] 2. Brought up things he/she knew you felt guilty about, to get you to do what he/she wanted you to do [1] 3. Used private Facebook messages, instant messages, text messages or email to gather information about you that he/she then used to blackmail you[1] 4. Acted sarcastically until you gave in to what he/she wanted[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Manipulation	
4B	Threatened to end the relationship as a way to get you to do what he/she wanted	5. Threatened to end the relationship as a way to get you to do what he/she wanted[3] 6. Threatened to spread rumors (in person or through social media) as a way to get you to do what he/she wanted[1] 7. Threatened to post pictures or videos of you on a social media site that you did not want posted (e.g., sexually explicit pictures), in order to get you to do what he/she wanted[1]
4C	Threatened to commit suicide as a way to get you to do what he/she wanted[a]	8. Threatened to commit suicide (whether or not you thought he/she would go through with it) as a way to get you to do what he/she wanted (e.g., not break up with him/her)[1] 9. Threatened to hurt him/herself (for example, by cutting) as a way to get you do what he/she wanted[1] 10. Said it was your fault that he/she wanted to hurt or kill him/herself, to make you feel bad or guilty[1]
	Public Humiliation	
5A	Threatened to reveal an embarrassing secret as a way to hurt or manipulate you	1. Threatened to reveal an embarrassing secret as a way to hurt or manipulate you[3] 2. Threatened to reveal an embarrassing secret you had told him/her in private as a way to hurt or manipulate you (e.g., a secret about a friend)[1]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Public Humiliation	
5B	Revealed important secrets to others that you had told him/her as a way to embarrass you	3. Revealed important secrets to others that you had told him/her as a way to humiliate you[3] 4. Started rumors about you at school or on the bus in order to give you a bad reputation[1] 5. Started rumors about your family in order to embarrass you[1] 6. Changed the information on your social media page(s) in order to embarrass you[1] 7. Posted detailed private information about your relationship in order to embarrass you[1] 8. Forwarded private messages or posted private conversations in order to humiliate you[1]
5C	Insulted or ridiculed you in front of others	9. Insulted or ridiculed you in front of others[3] 10. Started fights in public (e.g., in front of your friends) in order to embarrass you[1] 11. Posted something derogatory or offensive on a social media site, and then tagged you in it to humiliate you[1] 12. Liked a status that said derogatory or offensive things about you, in order to hurt you[1]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Verbal Abuse	
6A	Criticized and belittled you as a way to make you feel bad about yourself	1. Criticized you as a way to make you feel bad about yourself[1] 2. Sent you hurtful text messages to make you feel bad about yourself[1] 3. Brought up or did things that he/she knew were upsetting to you in order to hurt you[1] 4. Swore **at** you to make you feel bad about yourself[2] 5. Tried to make you feel like you were crazy or irrational[2] 6. Constantly pointed out your faults to make you feel bad about yourself[2] 7. Said you would never amount to anything to make you feel bad about yourself[2]
6B	Yelled and screamed as a way to intimidate you	8. Yelled and screamed as a way to intimidate you[3]
6C	Called you a derogatory name as a way to make you feel bad about yourself	9. Called you a derogatory name (e.g., for a girl, slut or whore; for a boy, pussy) as a way to make you feel bad about yourself[1] 10. Said things to hurt your feelings on purpose[2]
	Wound Through Sexuality	
7A	Criticized your physical looks or sexual performance as a way to humiliate you	1. Criticized your physical looks (e.g., called you fat or disgusting) in a sexual situation as a way to embarrass or humiliate you[1] 2. Criticized your sexual performance or sexual body parts as a way to embarrass or humiliate you (e.g., said you were bad in bed)[2] 3. Verbally bullied you in a sexual situation to make you feel inferior[1] 4. Told you that you were sexually and/or physically unattractive to him/her, in order to embarrass or humiliate you[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Wound Through Sexuality	
7B	Refused to have sex with you as a way of making you feel insecure or inadequate[a]	5. Made you do sexual things that you were not comfortable with, in order to embarrass or humiliate you[1] 6. Threatened to have sex with someone else if you did not do the sexual things he/she wanted[1] 7. Led you on to make you feel stupid or embarrassed[1] 8. Told you that if you didn't have sex with him/her, he/she wouldn't love you or would break up with you[1]
7C	Insisted you have sex with him/her in belittling or humiliating ways[a]	9. Made you do sexual things you were not comfortable with, in order to embarrass or humiliate you[1] 10. Refused to have sex with you because of your past sexual history, in order to make you feel humiliated[1] 11. Threatened to post sexually explicit videos of you if you didn't meet his/her sexual demands[1] 12. Made you feel like you were not good enough to have sex with (e.g., said you were dirty, called you a slut), to make you feel insecure[1] 13. Made you have sex in a way that you found demeaning, in order to humiliate you[1] 14. Told you that he/she was only with you for sex[1] 15. Gave you a hickey when you did not want one, or left other marks on your body, to show his/her ownership of you[2]
	Treat as Inferior	
8A	Tried to make you think he/she was more competent and intelligent than you as a way of making you feel inferior	1. Tried to make you think he/she was more competent and intelligent than you as a way of making you feel inferior[3] 2. Made negative comments to you about your intelligence, **not** while arguing, in order to make you feel inferior[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Treat as Inferior	
8B	Treated you as useless or stupid as a way to make you feel inferior	3. Treated you as useless or inadequate (for example, by comparing you to someone else) as a way to make you feel inferior[1] 4. Treated you as stupid (for example, by insulting your intelligence) as a way to make you feel inferior[1] 5. Continually told you or made you feel like you weren't good enough for him/her, to make you feel bad about yourself[1]
8C	Tried to demand obedience to orders that he/she gave as a way of establishing his/her authority over you[a]	6. Tried to demand obedience to orders that he/she gave as a way of establishing his/her authority over you[1] 7. Ordered you around to show his/her authority over you[2] 8. Gave you tasks as if he/she was 'in charge' of you[2] 9. Made you do something humiliating or degrading (e.g., begging for forgiveness, having to ask permission to do something) to show his/her authority over you[2]
	Hostile Environment	
9A	Intentionally turned a neutral interaction into an argument or disagreed with the purpose of creating conflict	1. Tried to turn everything you said into an argument or fight in order to create conflict[1] 2. Used an aggressive communication/interaction style, even when you were **not** arguing, in order to create conflict[1] 3. Looked through your private messages or Internet history in order to find things to argue about[1]
9B	Treated an argument as though he/she had to "drive you into the ground" and make you feel bad when making their points[a]	4. When arguing, constantly brought up things from the past to make you feel bad[1] 5. Confronted you about things you'd posted on social media site(s), or had sent in a private message, in order to start an argument[1]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Hostile Environment	
9C	Treated you with strong hatred and contempt[a]	6. Treated you with strong hatred and contempt when you tried to break up with him/her[1] 7. Treated you with strong hatred and contempt when you tried to speak your mind[1] 8. Blamed you for causing his/her violent or aggressive behavior[2] 9. Blamed you when he/she was upset about something, even when it had nothing to do with you[2] 10. Insisted that all the problems in the relationship were your fault[2]
	Monitoring	
10A	Tried to make you report on the details of where you went and what you did when you were not with him/her as a way to check on you	1. Tried to make you report on the details of where you went and what you did when you were not with him/her, as a way to check on you[3] 2. Expected you to be with him/her constantly, so that he/she could keep track of you/your actions[1] 3. Looked at your phone log or used *67 to see who you'd been talking to[1] 4. Made you call or text him/her constantly so that he/she could check up on what you were doing[1] 5. Monitored your time and made you account for where you were[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Monitoring	
10B	Listened in on phone conversations, read your email, or went through your belongings without your permission as a way to check on you	6. Constantly logged on to your social media page(s) (e.g., Twitter or Facebook) in order to check up on your activities[1] 7. Texted or called you constantly, when you did not want him/her to, in order to check up on your activities[1] 8. Demanded to look at your phone, or looked at your phone without your permission, to see who you were in contact with[1] 9. Showed up where he/she knew you would be (e.g., after-school job or friend's house), in order to check up on your activities[1] 10. Monitored your activities on social media site(s), and made you delete anything he/she didn't approve of, in order to control you[1] 11. Made you give him/her your social media or email password(s), so that he/she could monitor your private messages[1] 12. Hacked into your social media or email account(s) in order to read your private messages[1] 13. Went through your personal or private things (e.g., cell phone, bag etc.) without your permission[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Monitoring	
10C	Followed or had you followed by someone else as a way of checking up on your activities[a]	14. Had his/her friends watch or follow you as a way of checking up on your activities[1] 15. Wouldn't let you have any alone time or privacy in order to keep track of your activities[1] 16. Turned on a tracking device in your phone, without your knowledge or permission, in order to keep track of you[1] 17. Used your Facebook status or Twitter feed to show up where you were, when you didn't want him/her to[1]
	Wound Through Fidelity	
11A	Pointed out others as attractive as a way of making you feel uncomfortable	1. Pointed out others as attractive as a way of making you feel uncomfortable[3]
11B	Flirted with others in front of you as a way to make you jealous	2. Intentionally flirted with others in front of you in order to make you jealous[1]
11C	Implied he/she was cheating as a way to make you feel insecure and worried	3. Implied he/she was cheating as a way to make you feel insecure and worried[3] 4. Told you he/she was cheating on you in order to hurt your feelings[1] 5. Made sure you knew he/she was cheating on you (for example, by flirting on social media) in order to hurt you[1] 6. Told you he/she was sleeping with someone else in order to hurt you[1] 7. Did something unusual/mysterious/etc., which made you wonder if he/she was cheating on you[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Jealousy	
12A	Acted very upset because he/she felt jealous if you spoke to or looked at any person	1. Acted very upset because he/she felt jealous if you spoke to or looked at any other person[3] 2. Tried to forbid you from seeing or speaking to any ex-dating partners, because he/she was jealous[1] 3. Got angry with you because you talked to a particular person who he/she had forbid you from talking to[2] 4. Accused you of paying too much attention to someone or something else[2]
12B	Falsely accused you of trying to cheat on him/her, or actually cheating, as a way to restrict your behavior in order to prove you were not cheating	5. Falsely accused you of cheating, as a way to restrict your behavior in order to prove you were not cheating[1] 6. Used information he/she had obtained from your private messages or activities to accuse you of cheating[1] 7. Constantly accused you of cheating on him/her[2]
12C	Tried to prevent you from speaking to or looking at any person who could be a potential romantic partner for you	8. Tried to prevent you from speaking to or looking at any person who could be a potential romantic partner for you[3] 9. Acted upset if you spoke to or looked at any person who could be a potential romantic partner for you[2] 10. Did **not** allow you to speak to or look at any person who could be a potential romantic partner for you[2]
	Withheld Emotional/Physical	
13A	Ignored important holidays and events as a way to punish or hurt you[a]	1. Purposefully ignored days (for example, a birthday, anniversary or Valentine's Day) or events that were important to you, as a way to punish or hurt you[1] 2. Refused to do something that he/she knew was very important you, in order to punish or hurt you[2]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
Withheld Emotional/Physical		
13B	Refused to speak to you as a way to punish or hurt you	3. Refused to speak to you as a way to punish or hurt you[3] 4. Avoided you in public as a way to punish or hurt you[1] 5. Stood you up in order to embarrass or hurt you[1]
13C	Withheld physical or verbal affection as a way to punish or hurt you	6. Withheld physical or verbal affection as a way to punish or hurt you[3]
Control Personal Decisions		
14A	Acted very upset when he/she didn't get to make small decisions, such as what to watch on television or which restaurant to eat at[a]	1. Told you what to wear, and acted very upset if you didn't follow his/her instructions[1] 2. Expected you to ask for permission before you made certain decisions (e.g., what activities to do, what clothing to wear)[2]
14B	Tried to make personal choices that should have been left up to you (e.g., which clothes to wear, what you eat)[a]	3. Tried to make personal decisions (decisions you cared about) that should have been be left up to you (e.g., about who you are friends with)[1] 4. Made plans for you when you did not want him/her to (e.g., forced you to attend social events you didn't want to go to)[1]

(Appendix 5.A continues)

(Appendix 5.A continued)

Original MPAB Question (Follingstad, 2011)		Revised Question(s)
	Control Personal Decisions	
14C	Tried to make major decisions that affected you without consulting with you[a]	5. Tried to make big decisions for you, like what college to attend or whether to use birth control, that should have been left up to you[1] 6. Tried to make you stop attending school or other activities[1] 7. Tried to make you do things that would be bad for your future, so that you would need to rely on him/her (e.g., tried to tell you not to go to college)[1] 8. Tried to spend your money (for example, by taking your bank card)[1] 9. Made an important decision without consulting you because he/she thought it was their right as a male or a female to do so[2]
	Control Personal Decisions – Medical Care/Birth Control	
n/a	n/a	10. Kept you from using birth control or condoms, when you wanted to use them[1] 11. Kept you from getting birth control or condoms, when you wanted to get them[1] 12. Tampered with your form of birth control (for example, stole your pills, or poked holes in the condom)[1] 13. Hurt you physically, and then tried to stop you from going to the doctor about your injuries[1] 14. Kept you from using any form of birth control in order to get you pregnant (or, for males, got pregnant on purpose) when you did not want this to happen[1] 15. Stole, or made you give him/her, your prescription medications so that he/she could use them[1]

Note. MPAB questions taken from Follingstad (2011), Table 2, pp. 1203-1204.
[a]Item requiring adaptation
[1]Adapted MPAB item
[2]Additional item from other measures (Emelianchik-Key, 2010; Follingstad, Coyne, & Gambone, 2005; Leisen, 2000; Mac Neil, 2010; C. M. Murphy & Hoover,

1999; K. A. Murphy, Smith, & Xenos, 2012; Robert Wood Johnson Foundation, 2011; Shepard & Campbell, 1992; Thompson, Basile, Hertz, & Sitterle, 2006)
[3]Original MPAB item (Follingstad, 2011)

Appendix 5.B
PATS Survey

Psychological Aggression in Teens Scale (PATS) – Preliminary Survey

Section I: Dating History Screen

The following questions ask about your dating history. For these questions, dating means when you like a guy/girl and he/she likes you back, and other people know that you are together. This does not have to mean going on a formal date.

1. Which of the following best describes you?
   - *Heterosexual (straight)*
   - *Gay or lesbian*
   - *Bisexual*
   - *Not sure*

2. Please select the statement that best applies to you:
   - *I have not yet begun dating* ***[skip out of PATS survey]***
   - *I have begun dating*

3. What kinds of dating relationships have you had (either now or in the past)? Select all the answers that describe your dating history.
   - *Going out in male/female groups*
   - *Dating different people*
   - *Dating one person without any definite commitment*
   - *Dating one person exclusively*
   - *Engaged*

4. If you have ever been in a dating relationship or been going out with someone, please answer the following questions:
   - *At what age did you start going out/having girl/boyfriends?* _____
   - *How many girl/boyfriends have you had (not including childhood crushes)?* _____
   - *How many girl/boyfriends did you have/have you had in: (if a section is not applicable, please leave blank)*
     - Grade 9
       - Number of girl/boyfriends: _____
       - Longest relationship (weeks or months, please specify): _____
       - Shortest relationship (weeks or months, please specify): _____
     - Grade 10
       - Number of girl/boyfriends: _____
       - Longest relationship (weeks or months, please specify): _____
       - Shortest relationship (weeks or months, please specify): _____
     - Grade 11
       - Number of girl/boyfriends: _____
       - Longest relationship (weeks or months, please specify): _____
       - Shortest relationship (weeks or months, please specify): _____
     - Grade 12
       - Number of girl/boyfriends: _____
       - Longest relationship (weeks or months, please specify): _____
       - Shortest relationship (weeks or months, please specify): _____

The next set of questions asks you to think about things that might have happened to you in the past year. When answering these questions, we want you to think about one specific relationship.

So that we know which relationship you are thinking about, please answer the following questions:

5. What dating partner will you think about when answering the next set of questions?
   - *I will think of somebody that is my girl/boyfriend right now.* ***[skip to Q6]***
   - *I will think of a recent ex-girl/boyfriend* ***(within the past 3 months).*** ***[skip to Q15]***
   - *I will think of an ex-girl/boyfriend from* ***within the past year.*** ***[skip to Q15]***
   - *I have not had a dating relationship in the past year.* ***[skip out of PATS survey]***

For current partners:

6. How long have you been dating/going out with this person? (weeks or months, please specify)
   _____

7. How often do you see each other? Choose the best response.
   - *Every day at school*
   - *Every day at school and every day out of school*
   - *2-3 times per week*
   - *Once per week or less*

8. How much time do you spend alone together?
   _____ hours per day OR _____hours per week

9. What kinds of things do you do together? (please describe)
   _____

10. How often do you argue or disagree?
_____ times per day OR _____times per week

11. What kinds of things do you argue or disagree about? (please describe)
   _____

12. How old is he/she? _____

13. How important is this relationship to you? (circle the best response)
   - *Not very important*
   - *Somewhat important*
   - *Important*
   - *Very important*

14. Which of the following best describes your relationship with this person? (circle the best response)
   - *We go out in male/female groups*
   - *We date different people*
   - *We are dating each other, but without any definite commitment*
   - *We date each other exclusively*
   - *We are engaged*

{skip to Behavior questions}

For ex-partners:

15. How long did you date/go out together with this person? (weeks or months, please specify)
   _____

16. How often did you see each other? Choose the best response.
   - *Every day at school*
   - *Every day at school and every day out of school*
   - *2-3 times per week*
   - *Once per week or less*

17. How much time did you spend alone together?
    _____ hours per day OR _____hours per week

18. What kinds of things did you do together? (please describe)
    _____

19. When did you stop dating/going out together? _____

20. Why did you stop going out with him/her?
    _____

21. How often did you argue or disagree?
_____ times per day OR _____times per week

22. What kinds of things did you argue or disagree about? (please describe)
    _____

23. How old was he/she? _____

24. How important was this relationship to you? (circle the best response)
    - *Not very important*
    - *Somewhat important*
    - *Important*
    - *Very important*

25. Which of the following best described your relationship with this person? (circle the best response)
    - *We went out in male/female groups*
    - *We dated different people*
    - *We were dating each other, but without any definite commitment*
    - *We dated each other exclusively*
    - *We were engaged*

{skip to Behavior questions}

Section II: Behavioral Items

The following questions ask you about things that might have happened in your dating relationship. When answering these questions, think specifically about the relationship you identified above. Indicate experiences you have had in this relationship in the past year, by selecting Yes, No or Not Applicable. Answers to all questions are confidential, and you may stop participating at any time.

In the past year, this partner has:

1. Harmed or destroyed your personal things of value (e.g., pictures, clothing, cell phone) as a way to intimidate you
2. Threatened to break or destroy something that was important to you / that you valued as a way to intimidate you
3. Threatened to hurt one of your friends (physically or by bullying) in order to intimidate you
4. Threatened to spread rumors about a friend in order to intimidate you
5. Threatened to have his/her friends physically hurt someone you care about in order to intimidate you
6. Threatened to hurt something that means a lot to you (for example, a pet) in order to intimidate you
7. Drove recklessly when you were in the car in order to frighten you
8. Pressured you to drink alcohol, take drugs, or do something illegal, when you did not want to
9. Threw a temper tantrum (e.g., breaking/throwing objects, acting in a rage) as a way to frighten you
10. Threw things at you (e.g., a cell phone) as a way to frighten you
11. Continued to do something that scared you, even after you'd asked him/her to stop

12. Verbally threatened to physically harm you or made a gesture that seemed physically threatening as a way to frighten you
13. Physically grabbed you to show that he/she was more powerful, as a way to scare you
14. Said things to scare you (e.g., told you something "bad" would happen, threatened to commit suicide)
15. Threatened you with a knife, gun or other weapon
16. **Threatened** to kill you as a way to frighten you (whether or not you thought he/she would go through with it)
17. **Threatened** to kill you when you were planning to break up with him/her, or because he/she was scared of losing you
18. Acted rude toward, gossiped about, or told lies about your friends to discourage you from spending time with them?
19. Was constantly physically affectionate, when you did not want him/her to be, in order to show his/her ownership of you
20. Sent mean texts to your friends from your phone, or responded rudely to private Facebook or email messages from your account, in order to make people mad at you
21. Tried to keep you from socializing with your friends without him/her being present, in order to isolate you?
22. At school, tried to keep you away from everyone else, in order to isolate you?
23. Tried to control your social life (who you talk to, what you do on your phone or the internet) in order to isolate you
24. Tried to forbid you from seeing your friends, in order to isolate you?
25. Tried to forbid you from attending activities or events where he/she wouldn't be with you, in order to isolate you?
26. Forced you to skip school so that he/she could keep you away from other people
27. Pressured you to drop out of school so that he/she could isolate you
28. Kept you from attending your after-school activities, or pressured you to drop out of your after-school activities, in order to isolate you
29. Made you delete or de-activate a social media account (e.g., your Facebook account) in order to isolate you
30. Told you that you were only allowed to use social media (e.g, Facebook, Twitter) when he/she was present
31. Took your phone, and only let you use it when he/she was with you
32. Forbid you from using your phone when you were with him/her, so that you couldn't communicate with anyone else
33. Blocked people from your social media account(s), or deleted things from your account, when you did not want him/her to
34. Continued to act very upset (e.g., pouted, stayed angry, gave you the silent treatment) until you did what he/she wanted you to do
35. Brought up things he/she knew you felt guilty about, to get you to do what he/she wanted you to do
36. Used private Facebook messages, instant messages, text messages or email to gather information about you that he/she then used to blackmail you
37. Acted sarcastically until you gave in to what he/she wanted
38. Threatened to end the relationship as a way to get you to do what he/she wanted
39. Threatened to spread rumors (in person or through social media) as a way to get you to do what he/she wanted
40. Threatened to post pictures or videos of you on a social media site that you did not want posted (e.g., sexually explicit pictures), in order to get you to do what he/she wanted
41. Threatened to commit suicide (whether or not you thought he/she would go through with it) as a way to get you to do what he/she wanted (e.g., not break up with him/her)
42. Threatened to hurt him/herself (for example, by cutting) as a way to get you to do what he/she wanted
43. Said it was your fault that he/she want to hurt or kill him/herself, to make you feel bad or guilty
44. Threatened to reveal an embarrassing secret as a way to hurt or manipulate you
45. Threatened to reveal an embarrassing secret you had told him/her in private as a way to hurt or manipulate you (e.g., a secret about a friend)
46. Revealed important secrets to others that you had told him/her as a way to humiliate you
47. Started rumors about you at school or on the bus in order to give you a bad reputation
48. Started rumors about your family in order to embarrass you
49. Changed the information on your social media page(s) in order to embarrass you

50. Posted detailed private information about your relationship in order to embarrass you
51. Forwarded private messages or posted private conversations in order to humiliate you
52. Insulted or ridiculed you in front of others
53. Started fights in public (e.g., in front of your friends) in order to embarrass you
54. Posted something derogatory or offensive on a social media site, and then tagged you in it to humiliate you
55. Liked a status that said derogatory or offensive things about you, in order to hurt you
56. Criticized you as a way to make you feel bad about yourself
57. Sent you hurtful text messages to make you feel bad about yourself
58. Brought up or did things that he/she knew were upsetting to you in order to hurt you
59. Swore **at** you to make you feel bad about yourself
60. Tried to make you feel like you were crazy or irrational
61. Constantly pointed out your faults to make you feel bad about yourself
62. Said you would never amount to anything to make you feel bad about yourself
63. Yelled and screamed as a way to intimidate you
64. Called you a derogatory name (e.g., for a girl, slut or whore; for a boy, pussy) as a way to make you feel bad about yourself
65. Said things to hurt your feelings on purpose
66. Criticized your physical looks (e.g., called you fat or disgusting) in a sexual situation as a way to embarrass or humiliate you
67. Criticized your sexual performance or sexual body parts as a way to embarrass or humiliate you (e.g., said you were bad in bed)
68. Verbally bullied you in a sexual situation to make you feel inferior
69. Told you that you were sexually and/or physically unattractive to him/her, in order to embarrass or humiliate you
70. Made you do sexual things that you were not comfortable with, in order to embarrass or humiliate you
71. Threatened to have sex with someone else if you did not do the sexual things he/she wanted
72. Led you on to make you feel stupid or embarrassed
73. Told you that if you didn't have sex with him/her, he/she wouldn't love you or would break up with you
74. Made you do sexual things you were not comfortable with, in order to embarrass or humiliate you
75. Refused to have sex with you because of your past sexual history, in order to make you feel humiliated
76. Threatened to post sexually explicit videos of you if you didn't meet his/her sexual demands
77. Made you feel like you were not good enough to have sex with (e.g., said you were dirty, called you a slut), to make you feel insecure
78. Made you have sex in a way that you found demeaning, in order to humiliate you
79. Told you that he/she was only with you for sex
80. Gave you a hickey when you did not want one, or left other marks on your body, to show his/her ownership of you
81. Tried to make you think he/she was more competent and intelligent than you as a way of making you feel inferior
82. Made negative comments to you about your intelligence, **not** while arguing, in order to make you feel inferior
83. Treated you as useless or inadequate (for example, by comparing you to someone else) as a way to make you feel inferior
84. Treated you as stupid (for example, by insulting your intelligence) as a way to make you feel inferior
85. Continually told you or made you feel like you weren't good enough for him/her, to make you feel bad about yourself
86. Tried to demand obedience to orders that he/she gave as a way of establishing his/her authority over you
87. Ordered you around to show his/her authority over you
88. Gave you tasks as if he/she was 'in charge' of you
89. Made you do something humiliating or degrading (e.g., begging for forgiveness, having to ask permission to do something) to show his/her authority over you
90. Tried to turn everything you said into an argument or fight in order to create conflict
91. Used an aggressive communication/interaction style, even when you were **not** arguing, in order to create conflict
92. Looked through your private messages or Internet history in order to find things to argue about

93. When arguing, constantly brought up things from the past to make you feel bad
94. Confronted you about things you'd posted on social media site(s), or had sent in a private message, in order to start an argument
95. Treated you with strong hatred and contempt when you tried to break up with him/her
96. Treated you with strong hatred and contempt when you tried to speak your mind
97. Blamed you for causing his/her violent or aggressive behavior
98. Blamed you when he/she was upset about something, even when it had nothing to do with you
99. Insisted that all the problems in the relationship were your fault
100. Tried to make you report on the details of where you went and what you did when you were not with him/her, as a way to check on you
101. Expected you to be with him/her constantly, so that he/she could keep track of you/your actions
102. Looked at your phone log or used *67 to see who you'd been talking to
103. Made you call or text him/her constantly so that he/she could check up on what you were doing
104. Monitored your time and made you account for where you were
105. Constantly logged on to your social media page(s) (e.g., Twitter or Facebook) in order to check up on your activities
106. Texted or called you constantly, when you did not want him/her to, in order to check up on your activities
107. Demanded to look at your phone, or looked at your phone without your permission, to see who you were in contact with
108. Showed up where he/she knew you would be (e.g., after-school job or friend's house), in order to check up on your activities
109. Monitored your activities on social media site(s), and made you delete anything he/she didn't approve of, in order to control you
110. Made you give him/her your social media or email password(s), so that he/she could monitor your private messages
111. Hacked into your social media or email account(s) in order to read your private messages
112. Went through your personal or private things (e.g., cell phone, bag etc.) without your permission
113. Had his/her friends watch or follow you as a way of checking up on your activities
114. Wouldn't let you have any alone time or privacy in order to keep track of your activities
115. Turned on a tracking device in your phone, without your knowledge or permission, in order to keep track of you
116. Used your Facebook status or Twitter feed to show up where you were, when you didn't want him/her to
117. Pointed out others as attractive as a way of making you feel uncomfortable
118. Intentionally flirted with others in front of you in order to make you jealous
119. Implied he/she was cheating as a way to make you feel insecure and worried
120. Told you he/she was cheating on you in order to hurt your feelings
121. Made sure you knew he/she was cheating on you (for example, by flirting on social media) in order to hurt you
122. Told you he/she was sleeping with someone else in order to hurt you
123. Did something unusual/mysterious/etc., which made you wonder if he/she was cheating on you
124. Acted very upset because he/she felt jealous if you spoke to or looked at any other person
125. Tried to forbid you from seeing or speaking to any ex-dating partners, because he/she was jealous
126. Got angry with you because you talked to a particular person who he/she had forbid you from talking to
127. Accused you of paying too much attention to someone or something else
128. Falsely accused you of cheating, as a way to restrict your behavior in order to prove you were not cheating
129. Used information he/she had obtained from your private messages or activities to accuse you of cheating
130. Constantly accused you of cheating on him/her
131. Tried to prevent you from speaking to or looking at any person who could be a potential romantic partner for you
132. Acted upset if you spoke to or looked at any person who could be a potential romantic partner for you

133. Did **not** allow you to speak to or look at any person who could be a potential romantic partner for you
134. Purposefully ignored days (for example, a birthday, anniversary or Valentine's day) or events that were important to you, as a way to punish or hurt you
135. Refused to do something that he/she knew was very important to you, in order to punish or hurt you
136. Refused to speak to you as a way to punish or hurt you
137. Avoided you in public as a way to punish or hurt you
138. Stood you up in order to embarrass or hurt you
139. Withheld physical or verbal affection as a way to punish or hurt you
140. Told you what to wear, and acted very upset if you didn't follow his/her instructions
141. Expected you to ask for permission before you made certain decisions (e.g., what activities to do, what clothing to wear)
142. Tried to make personal decisions (decisions you cared about) that should have been be left up to you (e.g., about who you are friends with)
143. Made plans for you when you did not want him/her to (e.g., forced you to attend social events you didn't want to go to)
144. Tried to make big decisions for you, like what college to attend or whether to use birth control, that should have been left up to you
145. Tried to make you stop attending school or other activities
146. Tried to make you do things that would be bad for your future, so that you would need to rely on him/her (e.g., tried to tell you not to go to college)
147. Tried to spend your money (for example, by taking your bank card)
148. Made an important decision without consulting you because he/she thought it was their right as a male or a female to do so
149. Kept you from using birth control or condoms, when you wanted to use them
150. Kept you from getting birth control or condoms, when you wanted to get them
151. Tampered with your form of birth control (for example, stole your pills, or poked holes in the condom)
152. Hurt you physically, and then tried to stop you from going to the doctor about your injuries
153. Kept you from using any form of birth control in order to get you pregnant (or, for males, got pregnant on purpose) when you did not want this to happen
154. Stole, or made you give him/her, your prescription medications so that he/she could use them

---

Section III: Contextual Probes

If a question is endorsed *(note: context questions will not be asked on initial pilot)* –

A. How long ago did this experience occur? (select all that apply)
- *A week or less*
- *A month or less*
- *One to two months ago*
- *3 to 6 months ago*
- *More than 6 months ago*

B. Where were you when this experienced occurred? (select all that apply)
- *At my house*
- *At his/her house*
- *At a friend's house*
- *At school*
- *On school grounds/parking lot*
- *On the school bus or public transit*
- *In public (e.g., the mall, a friend's house)*
- *Other, please specify:* _____

C. Who was there when this experience occurred? (select all that apply)
- *We were alone*
- *My family members were there*
- *His/her family members were there*
- *My friends were there*
- *His/her friends were there*
- *Other, please describe: _____*

D. How long had you been seeing your boyfriend/girlfriend when this experience **first** occurred?
- *A week or less*
- *A month or less*
- *One to two months*
- *3 to 6 months*
- *More than 6 months*

E. How did you feel when this experience occurred? (select all that apply)
- *Upset*
- *Valued*
- *Confused*
- *Scared*
- *Appreciated*
- *Bothered*
- *Guilty*
- *Helpless*
- *Loved*
- *Embarrassed*
- *Respected*
- *Angry*
- *Special*
- *Threatened*
- *Worried*
- *Sad*
- *Controlled*

F. Approximately how many times did this experience happen in this relationship?
- *Rarely*
- *Sometimes*
- *Often*
- *Always or constantly*
- *I don't know*

G. In general, how much did this experience upset you?
(underlined portion is changed depending on item content)
- *Not at all*
- *A little bit*
- *A lot*
- *I don't know*

H. To what extent did you change your behavior as a result of this experience, in general?
(underlined portion is changed depending on item content)
- *Not at all*
- *A little bit*
- *A lot*
- *I don't know*

l. In general, do you think your boyfriend/girlfriend did this to be playful or as a joke?
   - *Yes*
   - *No*

Section IV: Overall Assessment

27. Overall, how psychologically abusive did/do you consider this partner?
   - Not psychologically abusive at all
   - A little bit psychologically abusive
   - Fairly psychologically abusive
   - Very psychologically abusive

Introductory questions were adapted from Wolfe et al. (2001). Behavior questions were adapted from Follingstad (2011). Context questions and overall assessment were adapted from Mac Neil (2010) and Follingstad and Edmundson (2010).